A comparative investigation of the attribution of desires and preferences

Rachel May Crosby

Newnham College

This dissertation is submitted for the degree of

Doctor of Philosophy

August 2019

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the relevant Degree Committee.

In Chapter 2, Ljerka Ostojić and Edward Legg designed Experiments 1 and 2 and were involved in discussions of the experimental design for Experiment 3. To calculate inter-observer reliability, Katharina Brecht acted as a second observer for Experiments 1 and 2, and Lovre Culina acted as a second observer for Experiment 3.

In Chapter 4, Ljerka Ostojić was involved in discussions of the experimental design for Experiments 1 and 2. In Experiment 1 the birds from Colony 2 were tested by Ljerka Ostojić.

In Chapter 5, the experimental design for Experiment 1 was based on an unpublished food sharing experiment (Ostojić, Legg, Mendl, & Clayton, 2014), and Ljerka Ostojić and Edward Legg were involved in discussions of the experimental design. The bird from Colony 2 was tested by Ljerka Ostojić. Experiment 2 was a replication of a published specific satiety caching study (Ostojić et al., 2017) and involved discussion with the original experimenter Ljerka Ostojić. The birds from Colony 2 were tested by Piero Amodio.

In Chapter 6, the experimental design for Experiment 1 was based on a pre-registration by Katharina Brecht (Brecht, Ostojić, Legg, & Clayton, 2016). For Experiment 1, Ben Farrar acted as a second observer for calculating inter-observer reliability.

Summary

Rachel Crosby

A comparative investigation of the attribution of desires and preferences

Although it is widely accepted that adult humans possess a 'theory of mind', debate has surrounded whether non-human animals may also be capable of attributing mental states such as knowledge, beliefs and desires to others. While researchers agree that animals are unlikely to possess a human-like theory of mind; theory of mind is now viewed as a continuum of social cognitive abilities and as such animals may possess limited elements of mental state attribution. A minimal form of theory of mind has been proposed by Apperly and Butterfill in human adults, which would allow rapid, efficient responses alongside a separate, slower, 'full blown' theory of mind. It has been suggested that this minimal system may be behind the limited theory of mind proposed in animals. Given that desires are representationally simpler than beliefs, desire attribution may be a good candidate for convergent minimal social cognitive abilities. In this thesis I therefore used Eurasian jays *(Garrulus glandarius)* as a comparator with human adults, due to existing evidence of Eurasian jays' sensitivity to their partner's desires. I focused on three key questions in my thesis:

1. Do adult humans have an implicit sensitivity to the desires and preferences of others?
2. Do Eurasian jays have a generalised sensitivity to others' desires and preferences?
3. Are non-human animals' responses to others' desires and preferences comparable to adult humans' implicit responses?

To address these questions I assessed adults' implicit sensitivity to others' desires in various scenarios, but did not find evidence of a consistent minimal system (Chapters 2, 3, 6). In addition, I investigated whether Eurasian jays' sensitivity to the desires of others may be applied outside of the food sharing context, as well as the flexibility of this sensitivity within food sharing (Chapters 4-6). Finally, I considered the similarities between the responses of Eurasian jays and humans and discuss the consequences of these findings for the hypothesis of a shared minimal system. I conclude by discussing the difficulties facing comparative cognition research and the possibility that theory of mind may be overestimated as a driver of social interactions in both humans and non-human animals.

Acknowledgements

I would like to thank Nicky Clayton for the opportunity to do my PhD in Cambridge, without her brilliant lab and welcoming culture I would not have been able to complete my thesis.

I am also indebted to Ljerka Ostojić, my amazing co-supervisor, who gave me guidance throughout my PhD, got me through many crises and continued to support me even after she left the lab. Thanks also go to Edward Legg for helping me with the North Aviary and discussing experiments in the tea room.

I would like to thank everyone who helped me out in Madingley; Katharina Brecht and Lovre Culina for their help with coding videos, Piero Amodio for testing birds for my experiment, and the technicians for always making sure the birds are in top condition. Finally, Ben Farrar has been my lifeline in Madingley with hours of crazy rants and cups of tea that kept me going through writing up my thesis. I will never forget Baldrick and Goosey.

Outside of the lab I want to thank my brilliant friend Becs for distracting me with board games and Kpop, as well as my family, particularly my parents, for putting up with my soapboxing about science. Last but not least I wouldn't have made it to the end without my incredible boyfriend Oli and all his patience and love over the last few years.

Publications

The findings from **Chapter 2** are published as a pre-print:

Crosby, R., Ostojić, L., Legg, E., & Clayton, N. (2018, December 23). No evidence found for anticipatory looking responses to specific satiety in adult humans. PsyArXiv, https://doi.org/10.31234/osf.io/c2rdu

The findings from **Chapter 6** have been submitted to Animal Behaviour and Cognition:

Crosby, R., Legg, E., Brecht, K., Mendel, M., Ostojić, L., & Clayton, N., Male Eurasian jays flexibly alter their food sharing in line with partners' choices.

Supplementary Information

In the interest of full transparency and reproducibility, the raw data and analysis scripts from all experiments, as well as examples of experimental materials can be found at: https://osf.io/suz2a/?view_only=500398f1349446a5967ea4b62293a3ad.

Contents

1. Chapter 1 - Introduction

The attribution of desires and preferences falls under the definition of 'theory of mind'; the ability to impute unobservable mental states, such as knowledge, beliefs or desires, to others (Premack & Woodruff, 1978). Although it is widely accepted that adult humans possess a theory of mind, intense debate has surrounded the question of whether non-human animals may also be capable of mental state attribution. While theory of mind has been implicated in studies of species ranging from dogs to dolphins, the current consensus is that it is unlikely that any non-human animals possess a human-like theory of mind. However, beyond this limited agreement there is still debate over whether non-human animals possess any form of theory of mind whatsoever (Heyes, 2015; Penn & Povinelli, 2007; Whiten, 2013). Researchers now view theory of mind as a continuum of different social cognitive abilities rather than an all-or-nothing trait (Call & Tomasello, 2008; Povinelli & Vonk, 2004). Thus, more recent research has focused on breaking down theory of mind to identify which components of mental state attribution non-human animals might possess. Here I give a brief review of the existing evidence for theory of mind in non-human animals as well as discussing some of the methodological and theoretical issues. I conclude by describing ways to address this issue and give an overview of the main aims and outcomes of my thesis.

1.1 Evidence for theory of mind in chimpanzees

Chimpanzees (*Pan troglodytes*) have been a main focus for research into mental state attribution ever since Premack and Woodruff coined the term 'theory of mind' (Premack & Woodruff, 1978). Despite initial enthusiasm, early research programmes obtained predominately negative results; finding that chimpanzees did not appear to understand false beliefs or respond to cues related to seeing (Call & Tomasello, 1999; Povinelli & Eddy, 1996). As such, it was concluded that chimpanzees did not possess a human-like theory of mind. However, later empirical studies yielded positive findings, leading to suggestions that chimpanzees may have an understanding of some aspects of theory of mind (Call & Tomasello, 2008). Nonetheless, these research programmes have often produced seemingly contrasting results and the interpretation of these findings has not been straightforward.

1.1.1 Goals and intentions

One aspect of theory of mind that has been investigated in chimpanzees is the understanding of others' goals and intentions. Initial studies found that chimpanzees did not appear to distinguish

between human actors who intentionally withheld food and those that accidentally failed to provide food (Povinelli, Perilloux, Reaux, & Bierschwale, 1998). However, in an alternative experiment chimpanzees and orangutans who were trained to search in marked boxes were more likely to search in a box that had been intentionally marked over a box that had been accidentally marked (Call & Tomasello, 1998). Furthermore, findings suggest that chimpanzees appear to produce more begging-related behaviours when faced with an experimenter unwilling to provide food versus an experimenter unable to provide food (Call, Hare, Carpenter, & Tomasello, 2004).

Finally, it has also been shown that chimpanzees are more likely to imitate the use of an unusual body part to perform an action (e.g. foot, head) when the actor's hands are free than when the actor's hands are occupied (D. Buttelmann, Carpenter, Call, & Tomasello, 2007). The authors described this as 'rational' imitation, because when the actor's hands are occupied they are constrained to perform the action with a different body part, whereas when their hands are free, use of a different body part may be a deliberate choice. Thus researchers argue that chimpanzees may have an understanding of intentional actions.

In addition, a number of studies have provided evidence that chimpanzees react appropriately to the goals of conspecifics (Call & Tomasello, 2008). For example, chimpanzees appear to help others by retrieving objects that they are unable to reach (Warneken, Hare, Melis, Hanus, & Tomasello, 2007; Warneken & Tomasello, 2006). However, it is not entirely clear from these experiments whether the chimpanzees' responses require an attribution of intentions or goals to others, or whether they are simply able to respond to observable differences in behaviour. In all of these studies the physical properties of the protagonist's actions differ between conditions, therefore it is not clear whether the chimpanzees are responding to the protagonist's *external goal* which is the state of affairs that the individual is trying to bring about (observable), or their *internal goal* which is the individual's mental representation of this outcome (unobservable) (Penn & Povinelli, 2013). Thus, chimpanzees' responses may be based directly on learnt or innate rules about external observable characteristics rather than any attribution of internal intentions or goals to others.

1.1.2 'Seeing' and perception

One of the most extensively utilised paradigms for investigating chimpanzees' theory of mind is the 'competitive feeding paradigm' (Hare, Call, Agnetta, & Tomasello, 2000). This involves a subordinate and a dominant chimpanzee at opposite ends of a training room with barriers in the centre of the room, behind which food can be hidden. Using this paradigm, subordinate

chimpanzees have been found to preferentially retrieve food when it is hidden from a dominant's view. In one experiment, in the uninformed condition food was placed behind one of two opaque occluders on the subordinate's side, either while the dominant's door was ajar (informed) or closed (uninformed) (Hare, Call, & Tomasello, 2001). In the misinformed condition, the food was placed behind one barrier while the dominant's door was ajar but then moved to the other barrier either while the dominant's door was ajar (informed) or closed (misinformed) (Hare et al., 2001). In both conditions, when given a head start, the subordinate was more likely to retrieve the food when the dominant had not seen it being placed (uninformed or misinformed). The authors therefore argue that these findings provide evidence for some understanding of 'seeing' on the behalf of the subordinate.

However, in a subsequent experiment within the same paper Hare et al. found negative results. When two pieces of food were placed behind the barriers, one that had been seen by the dominant and one that had not, the subordinate showed no preference for retrieving the unseen food over the seen food. The authors argue that this failure was due to a change in motivation caused by the addition of a second piece of food, rather than a lack of cognitive ability (Hare et al., 2001).

1.1.3 Knowledge

Claims have also been made that beyond understanding seeing, chimpanzees may also understand how seeing contributes to others' knowledge or ignorance. In Kaminski et al.'s experiment chimpanzees took turns choosing cups and receiving their contents, choosing either before or after a conspecific competitor (Kaminski, Call, & Tomasello, 2008). Two of the three cups were baited with food, the subject saw the baiting of both cups, and the competitor saw the baiting of one cup. The subjects were more likely to select the cup their competitor had not seen being baited when going second than when going first. The authors argue that this suggests the subject expected this cup to still contain food as they attributed ignorance to the competitor about this cup being baited. However, they showed a much lower performance on the task compared to 6-year-old children and adults.

Furthermore, if chimpanzees do have an understanding of knowing it may not be equivalent to the unified concept found in humans. There is conflicting information from studies on auditory information. Both chimpanzees and rhesus macaques have been shown to avoid being seen by humans while trying to steal food from them, and to choose quieter methods of stealing food (Flombaum & Santos, 2005; Hare, Call, & Tomasello, 2006; Melis, Call, & Tomasello, 2006; Santos, Nissen, & Ferrugia, 2006). However, in a competitive feeding paradigm experiment with

chimpanzees, when a dominant could hear but not see, subordinates did not preferentially take food that had been placed quietly over food that had been placed noisily (Bräuer, Call, & Tomasello, 2008). Thus, chimpanzees may take into account self-generated noise, but do not appear to take into account how other sources of noise may influence competitor's knowledge. As such, in contrast to the unified concept used by humans, any concept of 'knowing' in chimpanzees may be specific to certain modalities or scenarios (Apperly, 2010).

1.1.4 Beliefs

In addition, although there are some findings that chimpanzees may be sensitive to the knowledge of competitors, there is still limited evidence supporting a sensitivity to others' beliefs. Kaminski et al. conducted a false belief experiment using their competitive turn-taking paradigm (Kaminski et al., 2008). A high-value reward was placed in one of the three cups in view of both chimpanzees. The food was then either moved to a new location or picked up and replaced in the same location, either in view or out of view of the competitor. The competitor chose one of the cups out of view of the subject, and the subject was then given the opportunity to choose a cup or take a low-value reward. The subject was more likely to choose the cup with the high value reward when the competitor had not seen the final baiting. However, they did so regardless of whether the reward had been moved to a new location (competitor had a false belief) or remained in the same location (competitor had a true belief). Thus, the authors argue that chimpanzees may be able to respond successfully to the knowledge/ignorance of their competitor but do not take into account their beliefs.

In another false belief experiment chimpanzees were trained that an experimenter would bait the drawer indicated by a marker (O'Connell & Dunbar, 2003). In the test trials, before the baiting the marker was moved to point at a different drawer by an assistant. The experimenter then either double-checked the location of the marker after it was moved (true belief) or did not double-check the location of the marker (false belief). Thus, in the true belief condition the baited drawer should be indicated by the current location of the marker, whereas in the false belief condition the baited drawer should be at the original location of the marker. Similarly to the Kaminski et al. experiment, chimpanzees consistently chose the original location of the marker in the false belief trials; however they also did so (incorrectly) in the true belief trials. Thus, the chimpanzees may have been responding simply to the first location of the marker rather than responding to the belief of the experimenter.

Perhaps the most promising evidence for an understanding of beliefs in chimpanzees comes from a co-operative task, in which chimpanzees helped a human to retrieve an object (D.

Buttelmann, Buttelmann, Carpenter, Call, & Tomasello, 2017). The experimenter placed the object in one of two boxes. The assistant then moved the object to the other box, either while the experimenter was present (true belief) or absent (false belief). In both conditions the experimenter then tried to open the empty box. Following the logic of the authors, in the false belief condition the experimenter does not know the location of the object and thus is probably trying to retrieve the object from the wrong box. However, in the true belief condition the experimenter knows the location of the object and thus must be trying to open the empty box rather than trying to retrieve the object.

Chimpanzees were more likely to open the box containing the object in the false belief condition than the true belief condition. However, they did not follow the authors' prediction of preferentially opening the empty box in the true belief condition. The authors then conducted a second experiment involving an ignorance condition, in which the experimenter did not see the object being placed in its original location or being moved to the new location. The chimpanzees searched randomly in the ignorance condition, and the authors argued that this means that the chimpanzees' performance in the false belief condition could not be explained by simply attributing ignorance.

However, it is possible to explain the outcome of the above experiments based only on an understanding of the experimenters' knowledge/ignorance and external goals, rather than their beliefs. In the false belief condition, the experimenter is ignorant of the location of the object and attempts to open the wrong box, which is *consistent* with a goal of retrieving the object. In the true belief condition, the experimenter knows the location of the object but still attempts to open the wrong box, which is *inconsistent* with a goal of retrieving the object. Finally, in the ignorance condition, the experimenter never sees the object so they *cannot* have a goal of retrieving the object. Thus, the chimpanzees' random box opening in the true belief and ignorance conditions could be because they do not attribute the goal of obtaining the object to the experimenter. In contrast, in the false belief condition they could attribute this goal to the experimenter and thus act to help them get the object, similarly to previous studies of helping in chimpanzees mentioned above (Warneken et al., 2007; Warneken & Tomasello, 2006).

In summary, the current evidence suggests that, while chimpanzees may have some knowledge of the perceptions, goals and knowledge of others, there is no conclusive evidence that they are able to respond to others' beliefs. Furthermore, the understanding that chimpanzees do have of these mental states does not appear to be equivalent to a complete human-like theory of mind.

1.2 Evidence for theory of mind in corvids

While early animal ToM research predominately focussed on chimpanzees, there has been recent emphasis on corvids as candidates for advanced social cognition (Clayton, Dally, & Emery, 2007; Emery & Clayton, 2004). Although the evolutionary lines of mammals and birds separated about 300 million years ago, there has been increasing interest in the similarities between their cognitive skills; sparking claims of cognitive convergence between apes and corvids (Emery & Clayton, 2004; Seed, Emery, & Clayton, 2009; Van Horik, Clayton, & Emery, 2012). Specifically, corvids have been noted for their skill in tool use and manufacture (Bird & Emery, 2009; Weir, Chappell, & Kacelnik, 2002) their understanding of physical causality (Bird & Emery, 2010; Cheke, Bird, & Clayton, 2011; Davidson, Miller, Loissel, Cheke, & Clayton, 2017), and their potential for engaging in episodic-like cognition (Raby, Alexis, Dickinson, & Clayton, 2007).

There are two main viewpoints regarding the selective pressures driving the evolution of advanced cognition. One viewpoint, known as the 'social intelligence hypothesis', suggests that cognition may be an adaptation to the complex social interactions necessary for group living (Dunbar, 1998; Emery, Clayton, & Frith, 2007; Humphrey, 1976). While the other, known as the 'ecological intelligence hypothesis', suggests that the evolution of cognition might be driven by the need to track and predict the distribution of food, (Rosati, 2017). In corvids, both of these factors may have driven the evolution of cognitive abilities.

Corvids show huge variation in social systems across different species (Clayton & Emery, 2007). The most highly social species, such as rooks (*Corvus frugilegus*) and jackdaws (*Coloeus monedula*), live in large colonies year-round, maintaining strong and consistent social bonds with a number of individuals. At the other end of the spectrum are Eurasian jays (*Garrulus glandarius*), who are highly territorial and solitary, coming together with other individuals only in the breeding season, when they pair up and together aggressively defend their territory (Goodwin, 1951, 1956).

California scrub jays (*Aphelocoma californica*; formerly known as Western scrub jays) lie in the middle of this range, they are semi-territorial and will flock in small groups outside of breeding season (Clayton & Emery, 2007). Unlike their close relatives, Florida scrub jays (*Aphelocoma coerulescens)* and Mexican jays (*Aphelocoma wollweberi*), California scrub jays are not typically co-operative breeders. However, co-operative breeding is sometimes seen in the southern-most populations of California scrub jays in arid areas (Burt & Peterson, 1993; Dally, Emery, & Clayton, 2010; de Kort, Correia, Alexis, Dickinson, & Clayton, 2007).

Despite these variations in social systems across corvids, it has been suggested that the prevalence of sociality in corvids may indicate that the common ancestor of corvids was social,

and thus sociality may be a driving factor of advanced cognitive abilities across corvid species, regardless of their contemporary social structure (Clayton & Emery, 2007). However, it is unclear whether the relationship between social complexity and brain size, a crude measure that has been used to demonstrate the relationship between sociality and intelligence in primates (Dunbar, 1992), is also present in birds, as findings have been inconsistent (Beauchamp & Fernández-Juricic, 2004; Burish, Kueh, & Wang, 2004).

One proposed explanation for this is that there may be substantial flexibility in birds' social organisation or group sizes, both temporally throughout the year and spatially due to variation in environmental factors (Emery, Seed, von Bayern, & Clayton, 2007). Furthermore it has been suggested that the quality of relationships may be more important than the quantity of relationships for primarily monogamous corvids. The complex interactions found within long-term pair bonds have been suggested as a possible driver of advanced cognition, described as 'relationship intelligence', an offshoot of the social intelligence hypothesis (Emery, Seed, et al., 2007), see Chapter 7 for further discussion.

In addition to social pressures, ecological factors have also been investigated as potential drivers of cognitive evolution in corvids. One particularly well-explored area focuses on the cognitive demands of caching. Many corvids employ seasonal food caching; burying seeds and recovering them during the winter when food is scarce (Vander Wall, 1990). Similarly to sociality, corvids show wide variation in the extent to which they cache. For example, Eurasian jays are typically classified as 'intense' or 'specialised' cachers, California scrub jays are classified as 'moderate' cachers, and jackdaws are classified as 'non-cachers' (de Kort et al., 2007; Lucas, Brodin, De Kort, & Clayton, 2004). However phylogenetic reconstruction suggests that the common ancestor of corvids may have been a moderate cacher, suggesting that caching may have played a role in the evolution of cognition across the corvids (De Kort & Clayton, 2006).

It has been argued that caching particularly favours the evolution of spatial cognition, as these birds appear to be able to recall the locations of hundreds of cache sites, and return to them when food is scarce. In addition, cachers may benefit from remembering additional characteristics of the cache, such as how deep the item is buried, what type of food is stored and how long ago it was cached, particularly if the food item is perishable (Grodzinski & Clayton, 2010). It has therefore been hypothesised that caching may have driven the adaptive specialisation of spatial cognition, resulting in an increase in the volume of the hippocampus in caching birds (Krebs, 1990; Shettleworth, 1990).

Furthermore, the possibility of a link between social and ecological factors has also been discussed in the caching context. For any cacher, there is a risk of other individuals, both

conspecifics and heterospecifics, observing the cache being made and stealing the food before the cacher returns to retrieve it, an act known as 'pilfering'. Cachers living in social groups may therefore have to balance any benefits of group-living against the potential for increased pilfering of caches by conspecifics (Andersson & Krebs, 1978). Thus, a combination of caching and group-living may have driven the evolution of advanced social cognition by selecting for cachers that are able to attend to potential pilferers and employ strategies to assess and minimise the risk of pilfering (Grodzinski & Clayton, 2010).

1.2.1 **Cache protection strategies**

It has been suggested that corvids that do not live in groups may be less likely to be observed by a conspecific while caching in the wild, and there may therefore be less potential for pilfering of caches (De Kort, Emery, & Clayton, 2012). Studies have shown that sociality may affect the caching behaviour of corvids. For example, social Mexican jays are better than non-social Clark's nutcrackers (*Nucifraga columbiana*) at remembering the locations of caches made by conspecifics, despite Clark's nutcrackers being more intense cachers and more accurate in recovering their own caches (Bednekoff & Balda, 1996).

However, if the common ancestor of corvids lived in social groups and thus evolved cognitive strategies related to pilfering, parsimony would suggest that extant corvids would also possess these cognitive abilities to at least some extent. In line with this hypothesis, there is increasing evidence for social cognitive abilities in less social corvids. Moderately social ravens (*Corvus corax*), semi-territorial California scrub jays[1] and highly territorial Clark's nutcrackers and Eurasian jays have all been shown to employ 'cache protection strategies' that appear to limit the visual and acoustic information available to observing conspecifics during caching (Bugnyar & Heinrich, 2005; Clary & Kelly, 2011; Clayton et al., 2007; Shaw & Clayton, 2012).

1.2.1 Desires

Traditionally, the majority of research into theory of mind has been focused on the attribution of epistemic mental states such as knowledge and beliefs, with false belief tasks being emphasised as a gold standard measure of theory of mind (Wellman, Cross, & Watson, 2001; Wimmer & Perner, 1983). As a result, non-epistemic states such as desires have been somewhat neglected in non-human animal research. This is a questionable gap, given that our understanding and prediction of others' intentional actions is likely based around belief-desire reasoning (Wellman

[1] Henceforth referred to simply as scrub jays.

& Bartsch, 1988; Wellman & Woolley, 1990). That is, we anticipate that someone will perform an action if they *desire* an outcome and *believe* that the action will bring about that outcome.

In recent years, a series of experiments have been conducted to investigate Eurasian jays' sensitivity to the desires of conspecifics. It has been suggested that male jays may be able to respond to the desires of their partner during food sharing courtship behaviour (Ostojić, Legg, et al., 2016; Ostojić, Legg, Shaw, et al., 2014; Ostojić, Shaw, Cheke, & Clayton, 2013). In these studies, the desires of female jays were experimentally manipulated by inducing specific satiety; a sensory phenomenon in which continued sole consumption of a food leads to a reduced desire for that food (Hetherington & Rolls, 1996). After watching their female partner eat to satiety, males shared relatively less of the sated food. However, this was not the case when the female was pre-fed to satiety out of view of the male. This suggests that the difference in food sharing resulted from the male observing the female's eating, not only from differences in the female's behaviour during sharing (e.g. not taking the sated food from the male).

This finding was recently extended into the competitive caching context in a pooled study of scrub jays and Eurasian jays (Ostojić et al., 2017). Jays were given the opportunity to cache macadamia nuts and peanuts in front of an observer that had been pre-fed one of these two types of nuts. Cachers showed a greater preference for caching peanuts when the observer was pre-fed peanuts than when the observer was pre-fed macadamia nuts. The authors suggest that this was a cache protection strategy, as caching the nuts that the observer was sated on might reduce the likelihood of the caches being pilfered. This sensitivity to the observer's satiety during caching complements previous findings and suggests that jays may have a generalised ability to respond to others' desires rather than a specialised response specific to food sharing.

Desire attribution in Eurasian jays may therefore be a particularly promising area of investigation for theory of mind in corvids for a number of reasons. Firstly, food sharing and caching provide two ecologically valid contexts that allow the abilities of jays to be tested using natural behaviours. The fact that there is evidence across two contexts suggests that their sensitivity to desires is a generalised ability that may involve theory of mind, rather than an adaptive specialisation for food sharing. Secondly, the original food sharing effect has been demonstrated in a number of replications and conceptual expansions (Ostojić, Legg, et al., 2016; Ostojić, Legg, Mendl, et al., 2014; Ostojić, Legg, Shaw, et al., 2014; Ostojić et al., 2013), suggesting that this is a robust and reliable finding.

Finally, investigation of non-epistemic states such as desires has been limited to date, with the majority of non-human animal research being focused on goals, perceptions and beliefs as described above. One of the key questions in my thesis is therefore to what extent Eurasian jays'

sensitivity to others' desires is a flexible and generalisable cognitive ability. However, before attempting to answer this question, it is important to first consider the difficulties inherent in researching theory of mind in non-human animals.

On the one hand there exist the external constraints faced by all animal researchers, such as the availability of enough funding to allow a range of independent research groups with sufficient sample sizes to reach justifiable conclusions. Limited funding puts constraints on both the number of individuals a research group can support and they space available in which they can be housed. This may be a particular issue for corvids with their diverse social systems, as a group-living species will naturally occupy a much smaller space than the same number of individuals from a solitary species with non-overlapping territories.

For example, despite being naturally solitary, due to space constraints, the Eurasian jays at the University of Cambridge are housed in artificial colonies of up to sixteen individuals (except during breeding season when they are physically, but not visually, isolated into male-female pairs). This may increase the probability of jays gaining experience of their caches being pilfered, which may be an important factor for the development of cache protection strategies (Emery & Clayton, 2001). However, this does not resemble their natural social structure, meaning that the ecological relevance of the experimental scenarios may be diminished and that any conclusions drawn may not be relevant to their natural behaviour (see also Chapter 4).

On the other hand, there are the systemic incentive structures that affect how research is published and disseminated. Traditionally, high impact journals have focused on ground-breaking findings, favouring the publication of novel positive results, over negative results or replications. This can result in a biased literature with a high-rate of false positives, making it more difficult to construct a sound research program based on existing experimental paradigms (Meehl, 1967). These issues are compounded in corvid research by the external constraints described above, such as small sample sizes, a limited number of research groups and a lack of overlap in the species tested by these groups. If only a single research group is working with a given species, any independent replications must by necessity be conducted in a different species, and given the wide variation in social and ecological characteristics across corvids any failed replications could be due to differences in cognition between species or sites rather than a false positive original result (Farrar, Boeckle, & Clayton, 2020). I discuss these issues in greater depth in Chapter 7.

Finally, there lies the more fundamental question of how existing and future findings from animal cognition experiments should be interpreted; a question that must be considered before embarking on any research program within comparative cognition.

1.3 The 'argument by analogy' and the 'logical problem'

Despite strong convictions from certain researchers that both chimpanzees and corvids possess at least some form of theory of mind, there are equally vocal sceptics who argue that there is insufficient evidence to support this conclusion. Premack & Woodruff supplied the original and oft repeated definition of theory of mind; "an individual has a theory of mind if he imputes mental states to himself and others" (pg. 515) (Premack & Woodruff, 1978). However, this definition appeals to the subjective human experience of mental state attribution. The only way to apply this definition is through the 'argument by analogy'. Namely, that when humans and non-human animals produce similar behaviours we can reason by analogy that these similar behaviours are likely to be driven by similar psychological processes (Hume, 1739; Romanes, 1883).

This argument relies on two assumptions, (i) that similar behaviours in humans and non-human animals must have similar psychological causes, and (ii) that in humans these psychological causes can be reliably assessed by introspection (Povinelli, Bering, & Giambrone, 2000). Provided that these two assumptions hold we can use our own folk psychological understanding of our behaviours and apply these directly to scenarios in which non-human animals produce similar behaviours. However, there is good reason to be sceptical of both these assumptions. Firstly, similarity in behaviour may not always reflect similarity in psychology; and secondly, humans may be unreliable witnesses with regard to our own psychological processes (Penn & Povinelli, 2007; Povinelli et al., 2000; Povinelli & Giambrone, 1999). Thus, this verbal definition of theory of mind is ineffective because the argument by analogy does not represent a valid approach to determining the cognitive mechanisms underlying behaviour.

The disconnect between behaviour and cognition is emphasised by the 'logical problem', a formal argument for the limitations of current experimental paradigms used to assess theory of mind in non-human animals (Lurz, 2009; Povinelli & Vonk, 2004). Existing experiments testing theory of mind frequently assume that the animals must be attributing mental states in order to predict others' behaviour. However, as mental states must be inferred by the animal on the basis of observable cues, it is always possible that the animal could be responding directly to these observable cues rather than making the extra inferential step to unobservable mental states (Povinelli & Vonk, 2004).

An example of this problem in practice is given by the 'direct line of sight' argument, which challenges experiments that investigate an understanding of 'seeing'. In the chimpanzee competitive feeding paradigm (Hare et al., 2000, 2001), rather than attributing the mental state

'seeing' to the dominant, the subordinate may simply be responding to the observable cue of an unobstructed line of sight between the dominant's eyes and the food. Thus, the subordinate could act as if they were responding to the visual perspective of the dominant, while only ever responding to observable characteristics of the dominant's behaviour (orientation of head and eyes) and/or the environment (position of opaque barriers) (Whiten, 2013). Although these experiments show that chimpanzees have at least some sensitivity to others' perspectives, these paradigms do not necessarily require chimpanzees to attribute the mental state 'seeing' (Penn & Povinelli, 2007).

Similarly, equivalent arguments can be applied to corvid work. Cache protection strategies have been described in mentalistic terms; however it is also possible to explain these findings through reasoning only about observable characteristics. As Penn & Povinelli put it, "it suffices for the birds to associate specific competitors with specific cache sites and to reason in terms of the information they have observed from their own cognitive perspective" (pg. 736) (Penn & Povinelli, 2007). Consequently, sceptics argue that despite the wide range of studies purporting to provide positive evidence for mental state attribution in non-human animals, no currently applied test satisfactorily distinguishes theory of mind from alternative explanations (Heyes, 2015; Penn & Povinelli, 2007; Whiten, 2013).

Unsurprisingly, there have been a number of criticisms levelled at this sceptical position. Researchers in favour of a mentalistic account argue that alternative explanations are unparsimonious (Tomasello & Call, 2006). They suggest that responding to observable characteristics would quickly become inefficient as a new behavioural rule would be required for each novel scenario (Call & Tomasello, 2008; Whiten, 1996a). However, even with a full theory of mind an individual would still need to be able to categorise observable behaviours in order to infer the appropriate mental state. Thus, Povinelli & Vonk argue that in contrast to the assumption that theory of mind somehow simplifies the process, it instead adds an extra inferential step from behavioural abstractions to mental states (Povinelli & Vonk, 2004). While possessing a theory of mind may allow more efficient prediction and explanation of others' behaviour (Penn & Povinelli, 2007; Whiten, 1996a), it is still less parsimonious to assume that chimpanzees attribute mental states *in addition* to making behavioural inferences, rather than making behavioural inferences alone.

1.4 Addressing the 'logical problem'

When discussing theory of mind, researchers often divide animals' potential understanding into two broad categories 'mind-reading' and 'behaviour-reading'. Mind-reading covers the use of

theory of mind; attributing unobservable mental states to others' in order to predict, manipulate or explain their behaviour. In contrast, behaviour-reading covers all accounts in which the individual responds on the basis of observable cues without attributing unobservable mental states. However, this distinction between mind-reading and behaviour-reading is a false dichotomy. Although mental state attribution is frequently described as mind-reading, this shorthand is misleading as we are obviously unable to directly see the workings of others' minds. As Whiten puts it "mind-reading is not telepathy" (Whiten, 1996a); all inference regarding others' mental states must be based upon observable information. Thus, mind-reading appears to simply be an extended form of behaviour-reading.

In addition, the use of behaviour-reading as a catch-all term for alternatives to mind-reading has been criticised, as it can prevent researchers from considering lower-level explanations (Heyes, 2012, 2014b). Instead, behaviour-reading accounts often involve researchers appealing to their own 'common sense' to determine a conditional statement that the animal could be following e.g. *if* dominant has oriented to food *then* avoid (Heyes, 2015; Penn & Povinelli, 2007). As such, arguments have been made against behaviour-reading accounts as these 'behavioural rules' can be generated post-hoc by sceptics to explain any study's findings (Call & Tomasello, 2008; Halina, 2015). This has led to a stand-off in attempts to generate valid tests of theory of mind in animals. Researchers have already concluded that if non-human animals do possess a theory of mind it is almost certainly limited with respect to a full human-like theory of mind. However, paradigms are getting more intricate, such that any experiment researchers design to distinguish between mind-reading and behaviour-reading may be too complicated for animals to pass regardless of whether they possess theory of mind (van der Vaart & Hemelrijk, 2012). Furthermore, Povinelli et al.'s 'reinterpretation hypothesis' argues that the evolution of theory of mind in humans did not necessarily lead to the production of novel behaviours (Povinelli et al., 2000; Povinelli & Giambrone, 1999). Thus, the quest for a perfect experiment clearly demonstrating theory of mind in non-human animals may be futile. Instead, Heyes argues we should be looking for simpler precursors to theory of mind as this may allow progress to be made in understanding the evolution of human mind-reading (Heyes, 2015). Here I discuss some approaches suggested to address the logical problem and test for aspects theory of mind in non-human animals.

1.4.1 Self-other inference

Penn & Povinelli suggest that the only way to distinguish whether animals are capable of mind-reading is to develop a scenario in which the representation of mental states does some "unique causal work" beyond that achieved through the representation of observable characteristics

(pg.731) (Penn & Povinelli, 2007; Povinelli & Vonk, 2004). One potential paradigm repeatedly advocated by Povinelli is self-other inference, specifically variations of the 'goggles task' proposed by Heyes (Heyes, 1998). In this task, a subject is given the opportunity to familiarise themselves with two pairs of goggles, one of which can be seen through and the other which cannot be seen through. These goggles are externally indistinguishable from one another except for an arbitrary cue such as their colour. The subject is then given a task which involves another individual wearing these goggles. If the subject responds differently to the individual depending on which pair of goggles they are wearing, this may imply that the subject has inferred the individual's ability to see (or not see) from their own experience of the goggles.

Povinelli & Vonk argue that this paradigm has the potential to provide evidence for reasoning about mental states because the individual does not produce different behaviours when wearing the goggles; the only observable difference is the arbitrary colour cue (Povinelli & Vonk, 2004). Importantly, the scenario is also sufficiently novel that they believe the subject should not be able to react based on past experience. A number of studies have employed variations of self-other tasks with infants and children (Meltzoff & Brooks, 2008; Senju, Southgate, Snape, Leonard, & Csibra, 2011; Teufel, Alexis, Clayton, & Davis, 2010; Teufel, Clayton, & Russell, 2013). However, to date there is no study fully applying such a task with non-human animals.

A study involving a modified version of a self-other inference paradigm was conducted with chimpanzees (Karg, Schmelz, Call, & Tomasello, 2015). Here, the authors used gaze-following as a measure of whether chimpanzees recognised what a human experimenter could or could not see. The chimpanzees were given experience with 'face masks' that were either opaque or see-through and which differed in colour. In the test trials a human experimenter held one of the masks to her face and turned her head towards a camera. However, the chimpanzees' likelihood of looking towards the camera did not differ between the two masks.

A pilot study has also been conducted with two chimpanzees, involving a 'bucket' version of the task (Vonk & Povinelli, 2011). The chimpanzees were given experience with buckets that appeared opaque from outside but when placed over the head were either opaque or see-through and which differed in colour. The chimpanzees were then given the chance to beg for food from two experimenters wearing the buckets. Neither chimpanzee showed a preference for begging from the experimenter with the see-through bucket, thus failing to show any understanding of the experimenters' ability to see. However, these results have to be considered preliminary given the small number of subjects. It has also been suggested that this paradigm is too complex for chimpanzees given that in previous 'begging paradigms' they do not show a

consistent preference for experimenters without buckets on their head over those with buckets (Povinelli & Eddy, 1996; Reaux, Theall, & Povinelli, 1999; van der Vaart & Hemelrijk, 2012).

In addition to practical problems, there have also been questions raised over whether the goggles test is able to solve the logical problem in principle. One suggestion is that rather than inferring mental states from oneself to others, self-other inference tests may be passed by inferring from one's own *behaviours* (Andrews, 2005). If while wearing the bucket on their own head the chimpanzee finds that they are unable to complete tasks like grabbing objects and moving around they may make a behavioural link between wearing the bucket and not being able to do things. Thus, the chimpanzee may be able to pass the task by inferring that another individual wearing the same bucket may also have this behavioural incompetence, and therefore they may choose to beg from the other individual without attributing the mental state seeing.

Alternatively, it has been suggested that the goggles task could be solved by 'direct line of sight' (Low & Perner, 2012; Lurz, 2009). If there is an opaque barrier in the straight line between an individual's eyes and the object of interest, their line of sight is blocked. In contrast, a transparent barrier allows a direct line of sight. Heyes argues that this is not a solid basis for rejecting the goggles task as, while this argument is plausible, there is no evidence that chimpanzees conceive of opacity in this way (Heyes, 2015). However, Hare et al. have shown in their competitive feeding paradigm that chimpanzees do at least have some sensitivity to the difference in transparent and opaque barriers' effect on visual access; preferring food hidden behind an opaque barrier but not behind a transparent one (Hare et al., 2000).

1.4.2 Intervening variables

Another model of mind-reading that has been considered as an avenue for exploration in non-human animals is the intervening variables concept (Heyes, 2015; Whiten, 2013). Whiten avoids dichotomising mind-reading and behaviour-reading mechanisms, focusing instead on the question: when does behaviour-reading become mind-reading (Whiten, 1996a)? An individual (Z) must recognise any mental state in another individual (X) based on some observable characteristics of their behaviour or environment (see Figure 1.1). However, a range of different observable cues could conceivably relate to the same mental state, and similarly any state could lead to a number of appropriate responses. A mental state therefore represents an intervening variable that connects all the disparate behaviours that indicate this state (left) and result from this state (right).

Whiten argues that this concept of mental states as intervening variables provides the potential for effective investigation of theory of mind in non-human animals. If novel inputs (left) are

shown to drive the same adaptive outputs (right), or conversely if the same inputs can lead to novel adaptive outputs, this would indicate the presence of an intervening variable network as opposed to a series of learned "pairwise links" (Whiten, 2013). However, it may be empirically challenging to determine truly novel inputs (Heyes, 2015). For example, there may be observational similarities between behaviours indicating the same mental states, allowing stimulus generalisation. Alternatively, behaviours related to the same mental state may frequently occur together such that they are associated with one another; thus allowing the 'novel' inputs to produce the same outputs as the familiar inputs through mediated conditioning. As the 'novel' inputs must always be similar enough to other inputs to be recognisably related to the same mental state it may be impossible to fully account for these possibilities.

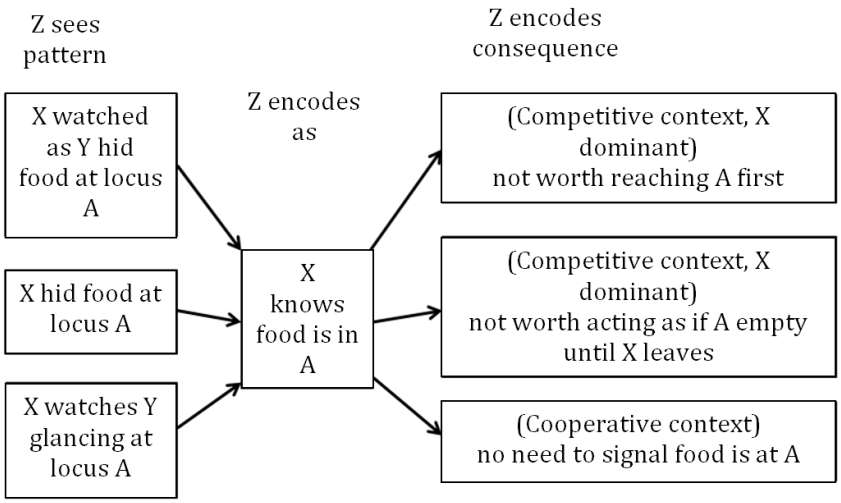

Figure 1.1 Diagram showing the mental state 'knowing' as an intervening variable, reproduced from (Whiten, 2013). The mental state can be inferred based on a number of different observable cues (left), and can lead to a number of potential behaviours (right) depending on the scenario.

1.4.3 Minimal theory of mind

A further clearly defined model that may allow for appropriate empirical exploration is the minimal system proposed by Apperly and Butterfill (Apperly, 2010; Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). Their key argument is that there exists a trade-off in theory of mind; some situations require deliberate, flexible analysis of behaviour, whereas other situations require rapid, consistent responses. They therefore propose a two systems model of mind-reading in adult humans, comprising a slow flexible system that represents mental states and a cognitively efficient rapid minimal system that represents the relations between agents, objects and locations as a proxy for mental states.

This appears to fit well with findings in infant theory of mind research. Traditional false belief tasks involve acting out a story and explicitly asking children to predict the actions of a character that holds a false belief about the location of an object[2] (Baron-Cohen, Leslie, & Frith, 1985; Wimmer & Perner, 1983). These tasks are passed around 4 years of age (Wellman & Liu, 2004). In contrast, more recently non-verbal measures have shown that infants may have some implicit sensitivity to others' false beliefs as early as the first year of life (Kovács, Téglás, & Endress, 2010; Onishi & Baillargeon, 2005; Scott & Baillargeon, 2017; Southgate & Vernetti, 2014; Surian, Caldi, & Sperber, 2007). Apperly and Butterfill suggest that this developmental dichotomy could be explained by a minimal system that contributes to implicit responses, alongside a later developing, flexible theory of mind that contributes to explicit responses. They outline four principles that might underlie such a minimal theory of mind system (Butterfill & Apperly, 2013): goal-directed action, encountering, registration, and an understanding that registration can act as a causal factor for behaviour (see Table 1.1).

The principles of encountering and registering can be defined entirely based on observable relations between agents, objects and locations (Heyes, 2015). As such, this minimal system can be re-characterised as a form of behaviour-reading based on a specific group of generalised behavioural rules. The minimal model may bridge the gap between behaviour-reading and mind-reading, breaking down the previously imposed false dichotomy between the two accounts. Apperly and Butterfill offer a re-description of existing experiments in light of their minimal model (Butterfill & Apperly, 2013). Cache protection strategies (see Section 2.1) have been characterised as evidence of visual perspective taking and an understanding of 'seeing', however they could instead result from a sensitivity to encountering (Clayton et al., 2007). Similarly, scrub jays' ability to track who saw what may be explained by registration rather than an understanding of others' knowledge states (Clayton et al., 2007; Dally, Emery, & Clayton, 2006).

In their discussion of the two systems approach, Apperly and Butterfill make a distinction between theory of mind abilities and theory of mind cognition (Butterfill & Apperly, 2013). A theory of mind ability allows an individual to track, respond to or influence the beliefs and perceptions of others, but does not necessarily require a representation of such states. In contrast, theory of mind cognition involves the representation of mental states, fitting more

[2] The most commonly described format is known as the Sally-Anne task: Sally places an object in location A, Sally leaves the scene, Anne moves the object to location B, and then the participating child is prompted to indicate where Sally will look for the object when she returns (Baron-Cohen et al., 1985). If the child has an understanding of false belief, they should recognise that Sally believes the object to be in location A and should look for it there, even though the object is now in location B.

closely with the traditional verbal definition of theory of mind. They argue that their minimal system qualifies as theory of mind because registration acts as an intervening variable and allows an individual to act in a way that tracks beliefs across different scenarios (Whiten, 1996b). However, they distinguish minimal from 'full blown' theory of mind as it does not meet the cognitive requirement of representing beliefs as such.

Table 1.1 Outline of the four principles that make up Apperly and Butterfill's minimal system (Butterfill & Apperly, 2013).

Principle	Explanation
(i) Goal-directed action	An understanding of the link between certain bodily movements and the goals to which they are directed. This principle may allow others' external goals to be tracked without representing any internal intentional states (see Section 1.1).
(ii) Encountering	An understanding that agents will only act towards objects within their field. The scope of the agent's field can be determined by observable characteristics such as the agent's orientation or the presence of opaque barriers. This principle may allow others' perceptions to be tracked without representing 'seeing' as a mental state (see Section 1.2).
(iii) Registration	An understanding of the relation between an agent, an object and a location; "an individual registers an object at a location if and only if she most recently encountered it at that location" (pg. 617) (Butterfill & Apperly, 2013). In order to act towards an object an agent must have registered that object in that location. This principle may allow the manipulation of others' knowledge about the location of objects without representing knowledge as a mental state (see Section 1.3).
(iv) Causality	An understanding that registration can influence the actions taken by an agent. An agent will act as if an object is located in the location they registered it at, even if this registration is now incorrect. This principle may allow registration to act as a proxy for belief, producing appropriate responses in false belief tasks without requiring the representation of beliefs as mental states (e.g. infant implicit false belief understanding, see Section 4.5).

1.4.4 Evidence for minimal theory of mind in human adults

A key element of Apperly and Butterfill's argument is that the minimal system is maintained into adulthood, and acts alongside an explicit full blown theory of mind. The explicit system allows for flexible consideration of others' behaviour, while the minimal system allows the efficient computation necessary for rapid responses. In support of this theory, Apperly and Butterfill cite contradictory findings regarding the cognitive demands and automaticity of theory of mind in human adults. There is a large body of evidence suggesting that reasoning about others' mental states is dependent upon executive control, attention and working memory (Apperly, Samson, & Humphreys, 2009; Bull, Phillips, & Conway, 2008; German & Hehman, 2006; Hughes, 1998; Lin, Keysar, & Epley, 2010; McKinnon & Moscovitch, 2007). This is complemented by evidence that theory of mind is non-automatic (Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006; Back & Apperly, 2010) and incurs processing costs (Apperly, Back, Samson, & France, 2008).

However, there are also contrasting findings that suggest that theory of mind processing can occur spontaneously, even when this is not necessary for the task (Kovács et al., 2010; Samson, Apperly, Braithwaite, Andrews, & Scott, 2010)[3]. Individuals appear to automatically represent the perceptions of others in a way that interferes with the processing of their own perceptions, a phenomenon described as 'altercentric interference' (Samson et al., 2010). Furthermore, a dual-task study found that level-1 perspective taking may occur independently of executive function, with executive function only required for additional processes such as perspective selection (Qureshi et al., 2010). Apperly and Butterfill argue that these seemingly contrasting results can only be reconciled if two systems exist within theory of mind (Butterfill & Apperly, 2013). The automatic processing of perspectives in adults and infants is accounted for by the minimal system, while full blown theory of mind improves gradually, in line with the development of executive function and language in children (Astington & Baird, 2005; Austin, Groppe, & Elsner, 2014; Hughes & Ensor, 2007; Sabbagh, Xu, Carlson, Moses, & Lee, 2006).

[3] It must be noted that some important criticisms have been made regarding these studies, suggesting that they do not tap into an implicit theory of mind as claimed by the authors. The Samson et al. paper utilised a dot perspective task and found that participants' speed and accuracy was affected by the perspective of an avatar (Samson et al., 2010). However, Heyes criticises their conclusion, arguing that this difference could be due to attentional factors rather than the attribution of perspective (Heyes, 2014b). This hypothesis was supported by a later study reinvestigating the task (Santiesteban, Catmur, Hopkins, Bird, & Heyes, 2014). Furthermore, the adult experiment in the Kovács et al. paper suggested that participants were automatically representing an agent's beliefs, resulting in faster reaction times (Kovács et al., 2010). However, a later study suggested that this may have been an artefact of the experimental paradigm (J. Phillips et al., 2015).

In order to achieve cognitive efficiency the minimal system must trade off flexibility for speed, and as a result the minimal system must be limited relative to full, flexible theory of mind. The minimal system is therefore characterised in adults by certain 'signature limits'. One such signature limit suggested by Apperly and Butterfill is that while the minimal system may be able to track the relation between agents and objects i.e. *whether* an object is perceived (level-1 perspective taking) it may not be able to track *how* an object is perceived (level-2 perspective taking).

This signature limit has been demonstrated in adults in an anticipatory looking experiment (Low & Watts, 2013). In familiarisation trials an actor was shown to consistently choose blue toys over red toys. The test trial involved a toy which appeared red from one side and blue from the other. The dual-sided toy was shown moving from the left to the right box, with the blue side facing the actor and the red side facing the participant. The toy then span around out of view of the actor and moved back to the left box with the red side facing the actor and the blue side facing the participant. Following the authors' logic, the actor should believe that a blue toy was still in the right box and a different red toy had moved into the left box. Adults correctly verbally predicted that the actor would search in the box on the right (95%). However, their anticipatory looking showed the opposite pattern, with only 25% of adults looking towards the right box. The authors argue that this lack of second order perspective taking in the automatic eye movement response is a signature limit of their minimal system and thus this experiment provides evidence in favour of their theory.

1.4.5 Applying minimal theory of mind to desires

However, the question that has yet to be answered is how this minimal system fits into a traditional belief-desire psychology (Wellman & Bartsch, 1988). A number of systematic studies and meta-analyses assessing children's theory of mind using verbal measures have found a consistent hierarchy in the development of certain mental state concepts (Wellman & Liu, 2004; Wellman & Woolley, 1990). Children typically appear to first understand others' perceptions and desires, followed later by an understanding of beliefs, and finally false beliefs around the age of four. It has been suggested that desires may be simpler to conceive of than beliefs, as desires are not inherently representational (Wellman & Woolley, 1990). Whereas beliefs are attitudes about a proposition ('he believes *there is an apple'*), desires can be attitudes about actual objects ('he wants *an apple*'). The former requires an internal representation of the external object, while the latter requires only a longing for the external object. Thus, attributing beliefs inherently

requires meta-representation, while this is not strictly necessary for a representation of others' desires.

However, despite consistent conclusions based on children's verbal responses, as described above a large number of recent studies have claimed to show an earlier understanding of false belief in infants using spontaneous 'implicit' measures. In contrast to verbal reasoning, looking time measures such as anticipatory looking and violation of expectation have been used to provide evidence for an understanding of false belief in infants in their first year of life (Scott & Baillargeon, 2017). Proponents argue that these measures are an indication of early theory of mind competence (Baillargeon, Scott, & He, 2010; Carruthers, 2013). Although sceptics argue that these spontaneous responses may arise due to other factors such as perceptual novelty (Heyes, 2014a; Perner & Ruffman, 2005)[4].

As the field currently stands, it is clear that an *explicit* verbal conception of desires arises earlier than an explicit understanding of beliefs. However, it is possible that this pattern of development for explicit understanding may also be mirrored at the *implicit* level, with an implicit response to desires emerging even earlier than an implicit sensitivity to beliefs (Steglich-Petersen & Michael, 2015). As outlined in Table 1.1 above, the first principle of Apperly and Butterfill's minimal system is goal-directed action. There is already extensive evidence suggesting that infants have some form of sensitivity to goal-directed action as early as 3 months of age (Sommerville, Woodward, & Needham, 2005; Woodward, 2009). The majority of these studies use violation of expectation measures where infants' 'surprise' is assessed by measuring the amount of time they spend watching an event. The standard example of a goal-attribution experiment is to show infants a video of an actor reaching for and grasping an object repeatedly and then show videos of the actor reaching either for the same object or a novel object (Woodward, 1998). If the infants spend a longer time looking at the video showing the actor reaching for the novel object this is taken to suggest that the infant found the event surprising and that their expectation of what would happen was violated.

This format has been used in a number of different studies, with findings suggesting that infants show flexible attribution of goals. Researchers have argued that, when attributing goals, infants appear to be sensitive to the identity and features of an agent (Buresh & Woodward, 2007; Henderson & Woodward, 2012; Luo & Baillargeon, 2005; Shimizu & Johnson, 2004), the

[4] There have also been a number of recent failed replications relating to implicit false belief experiments (e.g. Kulke, von Duhn, Schneider, & Rakoczy, 2018; Poulin-Dubois et al., 2018; Powell, Hobbs, Bardis, Carey, & Saxe, 2018) (see Chapter 7).

efficiency or effectiveness of the agent's actions (Biro, Verschoor, & Coenen, 2011; Woodward, 1999; Woodward & Sommerville, 2000), and the attention of the agent (Johnson, Ok, & Luo, 2007; A. T. Phillips, Wellman, & Spelke, 2002).

However, there is currently no consensus on whether the infants are attributing mental states (i.e. intentions or internal goals) to agents (Woodward, 2009), or responding only to observable aspects of the experiment such as actions or external goals (Gergely & Csibra, 2003; Ruffman, Taumoepeau, & Perkins, 2012) (see section 1.1). A third stance, is that the infants' looking responses do not result from an understanding of goals at all but are simply a consequence of low-level novelty (Heyes, 2014a), although this stance has been criticised by goal-directed action researchers, c.f. (Scott & Baillargeon, 2014). Apperly and Butterfill's characterisation of goal-directed action is most similar to the second stance, strictly referencing the relations between actions and external goals without any representation of mental states as such.

However, violation of expectation experiments appear to show that infants are not only sensitive to agent's goals, but also have some understanding of how preferences are formed. For example, infants seem to show no expectation that an agent will choose object A over object B when they have only previously seen the protagonist reaching for object A when object B was not present/visible (Luo, 2011; Luo & Baillargeon, 2007) (see also Chapter 6). To explain this, Apperly and Butterfill suggest that goal-directed action could be combined with registration to track preferences (Butterfill & Apperly, 2016). They characterise a preference as a probabilistic relationship between goals and an agent. If an agent performs an action to achieve a goal A as opposed to goal B, the probability of them subsequently performing an action directed to goal A rather than goal B increases i.e. they *prefer* goal A. Hence, a preference can be tracked probabilistically without representing any of the mental states or intentions that might underlie this preference (Ruffman et al., 2012).

Furthermore, in addition to explaining goal-directed action and simple preferences, Apperly and Butterfill have also suggested that their minimal system could be extended to include a more general tracking of desires. They state that "further notions, such as a relational proxy for tracking desires, could also be added. These and other modifications would enable hypotheses about minimal theory of mind cognition to explain a wider range of theory of mind abilities" (pg.620) (Butterfill & Apperly, 2013).

There have been some studies that have investigated the possibility of an implicit sensitivity to diverse desires i.e. the understanding that different individuals can have different desires. Repacholi and Gopnik tested children's understanding of diverse desires using a non-verbal intentional action paradigm (Repacholi & Gopnik, 1997) (see also Chapter 4). The experimenter

tasted each food and produced either a positive or negative response – an exclamation 'Mmm!' or 'Eww!' combined with the appropriate facial expression. The child's own preference was recorded and the experimenter's preferred food either matched or contrasted with the child's preferred food. When the experimenter requested food from the child, 14-month-olds provided the experimenter with the food they preferred themselves, while 18-month-olds provided the experimenter with the food that the experimenter had reacted positively to. The researchers suggest that this indicates that children have an understanding of diverse desires by 18-months-old.[5]

A similar result was later found in a violation of expectation study (Yott & Poulin-Dubois, 2016). Here, 18-month-olds were shown an experimenter reacting to two foods, one positively and one negatively. The children looked for longer when a second experimenter gave the first experimenter the negative food then when they gave the first experimenter the positive food. The authors suggest that this indicates that the children attributed a desire to the first experimenter for the food they reacted to positively, and they were therefore surprised when the second experimenter gave them the undesirable food.

However, it is not clear why attributing a desire to the first experimenter would cause the children to be surprised about the actions of the second experimenter. The assumption is that the infant expects the second experimenter to give the desired food to the first experimenter. This expectation would require either that the infant attributes an intention to the second experimenter to fulfil the first experimenter's desire, or that the infant has previous experience of how people act in similar situations. The latter option appears more likely, however this opens up the possibility that the infant did not attribute a desire to the first experimenter at all, instead responding entirely on the basis of this previous experience (Perner & Ruffman, 2005). It is therefore not yet clear from these experiments alone whether children have an implicit sensitivity to the desires of others.

Further research is necessary to determine whether the minimal system is capable of contributing to a wide range of theory of mind abilities or is limited to the specific cases of goal-

[5]A similar study investigated chimpanzee's understanding of a human experimenter's facial expressions (D. Buttelmann, Call, & Tomasello, 2009). The study involved three experiments using the same subjects, the subjects were only rewarded when they made the correct choice, the subjects were given multiple trials in each experiment, and the predicted effect was found only in the final study. The chimpanzees may therefore have improved over time through trial and error learning rather than attributing desires to the experimenter or making any predictions about the experimenter's behaviour based on their facial expressions.

attribution, perspective taking and beliefs that have been studied predominately to date. Thus, the second key question I consider in my thesis is whether human adults might have an implicit sensitivity to desires and preferences. I focus on adult humans as this allow more versatility of experimental design, and to my knowledge there have not yet been any studies investigating adults' implicit responses to others' desires.

1.4.6 Evidence for minimal theory of mind in non-human animals

As well as considering the presence of a minimal system in human adults, Apperly and Butterfill also suggest this mechanism may represent a fundamental basis for theory of mind shared with non-human animals. In reference to theory of mind in infants and animals, they state that "there is evidence that these abilities are limited and limited in similar ways" (pg. 958) (Apperly & Butterfill, 2009). Evidence in favour of this possibility is provided by anticipatory looking experiments with great apes. In a test of false belief understanding a human competitor saw one of two containers being baited with a reward (Krachun, Carpenter, Call, & Tomasello, 2009). The reward was then moved either in view (true belief) or out of the view (false belief) of the competitor. The competitor then tried, and failed, to reach for the container they believed contained the reward. The apes chose the indicated container across all trials, regardless of whether the competitor had a false belief about the location of the reward. However, the authors noted that the apes appeared to look more often at the non-indicated container in the false beliefs trial than the true belief trials.

This finding was then expanded upon in a study directly accessing the anticipatory looking of great apes in a false belief task (Krupenye, Kano, Hirata, Call, & Tomasello, 2016). The study involved showing apes videos of two modified false belief tasks. In the first of these tasks a human actor searched for an individual dressed like an ape in one of two haystacks. The human actor always saw the original location of the character, however, the character then moved between locations either while the actor was watching (true belief) or while the actor was absent (false belief). The character then moved away from the final location while the actor was absent, so that the apes would not simply look to the last location of the character. Finally, the experimenters measured the anticipatory eye movements of the apes when the actor re-entered the scene. The second experiment was similar to this, except that instead of searching for the ape-like character, the human actor searched for an object hidden by the character in one of two boxes.

The authors found that in Experiment 1, out of the apes that made an anticipatory look (30/40), although more than half looked first towards the location that the actor believed the character

to be in (20/30) this was not more than expected by chance. In contrast, in Experiment 2 the apes that made an anticipatory look (22/30) were more likely to look first towards the location the actor believed the object to be in than expected by chance (17/22). The authors then pooled the data across the two experiments and concluded that the apes "accurately anticipated the goal-directed behaviour of an agent who held a false belief" (pg. 113) (Krupenye et al., 2016). They additionally argued that given the lack of success shown by apes in tasks explicitly measuring false belief this finding may result from an implicit understanding of belief. As a result, some researchers consider this finding to be evidence of an implicit or minimal system of theory of mind shared between humans and non-human animals (Bugnyar, 2017). Hence, my final key question in this thesis is whether there may be similarities between the desire sensitivity of Eurasian jays and the implicit sensitivity of adult humans that could potentially stem from a shared minimal system.

1.5 Overview of thesis

In summary, while a range of studies have been suggested to provide evidence for theory of mind in non-human animals (Call & Tomasello, 2008), sceptics argue that their results might be equally compatible with alternative non-mentalistic explanations (Penn & Povinelli, 2007). However, one promising area for further research may come from studies showing Eurasian jays' sensitivity to desires in the contexts of both food sharing and food caching. Despite the validity of the logical problem, progress can still be made in the search for the mechanisms underlying mind-reading in both humans and non-human animals. Apperly and Butterfill's minimal theory of mind system may be a strong candidate for a limited form of theory of mind in adult humans and infants (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). Furthermore, there is tentative evidence that this system may be shared by humans and non-human animals (Bugnyar, 2017; Krupenye et al., 2016). In this thesis I investigate the possibility of a shared minimal theory of mind system contributing both to a response to desires in Eurasian jays and an implicit response to desires in human adults. Given the existing evidence for sensitivity to desires in Eurasian jays and given that desires may be representationally simpler than beliefs, this investigation could reveal the presence of convergent minimal social cognitive abilities.

I describe my findings in five experimental chapters, in which I address three key questions:

1. Do adult humans have an implicit sensitivity to the desires and preferences of others (see Section 4.5)?

2. Do Eurasian jays have a generalised sensitivity to others' desires and preferences (see Section 2.2)?
3. Are non-human animals' responses to others' desires and preferences comparable to adult humans' implicit responses (see Section 4.6)?

The current evidence for Eurasian jays' sensitivity to others' desires comes from experiments involving manipulation of their partners' desires via specific satiety. In **Chapter 2**, I therefore started by exploring whether adult humans may also have an implicit sensitivity to satiety. I employed anticipatory looking as a measure of implicit understanding, which has been used extensively in researching both infants' and adults' implicit responses to false beliefs. Participants watched a protagonist[6] eat food to satiety and were then prompted that a protagonist was about to make a choice between two foods. Participants did not demonstrate any directed anticipatory looking unless the protagonist provided an additional verbal cue regarding their satiety. This is an interesting result given the existing evidence for Eurasian jays' sensitivity to specific satiety in the food sharing context. However, specific satiety is not the only strategy that can be used to assess an implicit understanding of others' desires.

In **Chapter 3**, I extended my investigation by assessing whether human adults are capable of implicitly determining agents' preferences[7] through transitive inference. It has been demonstrated that humans can spontaneously attribute preferences to another individual or agent when observing their choices between objects. Research with infants using violation of expectation paradigms has suggested that infants' looking times reflect a spontaneous understanding of the transitivity of others' preferences e.g. if A>B and B>C, then A>C. Given the current uncertainty over the effectiveness of violation of expectation measures I used an intentional action measure of adults' implicit understanding of the transitivity of an agent's preferences. I found some evidence for an implicit response to transitive inference in adult humans, although it is unclear to what extent this response was reliant on logical reasoning about the hierarchy.

To complement the research in adult humans, I also add to existing research regarding Eurasian jays' understanding of others' desires. In **Chapter 4**, I explored Eurasian jays' co-operative social cognitive abilities outside the context of beak-to-beak food sharing by designing novel paradigms to test jays' sensitivity to the specific satiety of their partner. I investigated whether

[6] Throughout this thesis, I use the term 'protagonist' when referring to a human actor and I use the term 'agent' when referring to an animated character.

[7] I use the term 'preference' specifically when a choice is being made between two objects.

the looking behaviour of jays showed anticipation of their partner's presence. This experiment involved the jays looking through one of two peepholes; however validation of the paradigm failed due to a low frequency of looking through the peepholes and bias towards looking through the peephole on the left side. In a further paradigm, jays were trained to pull strings to provide different types of food to their partner's compartment. The jays passed the training stages, however they reached extinction on the behaviour due to a lack of response from their partner. These findings suggest that beak-to-beak food sharing may be a specifically useful paradigm for social cognitive investigations as birds show lower motivation to interact in other scenarios. I therefore conducted a pilot study investigating the inherent motivation underlying food sharing behaviours, and discuss the possibilities for a future experiment.

In Chapter 4, I was unable to find a paradigm that could be used to investigate Eurasian jays' sensitivity to their partner's desires outside of beak-to-beak food sharing. However there is an existing paradigm that has been used to show Eurasian jays' sensitivity to others' desires in a comparative caching context. In **Chapter 5**, I therefore investigated the desire attribution abilities of Eurasian jays in a caching scenario by manipulating the observable cues available to the cacher. Unexpectedly, the cachers' caching pattern did not differ with the observers' satiety. This was a surprising result as the experiment included a conceptual replication of previous findings. To explore this outcome, I subsequently attempted to replicate the original study that identified the effect (Ostojić et al., 2017). I failed to replicate the original finding from either experiment of the study, suggesting that if cachers are sensitive to an observer's specific satiety this may not be a robust or reliable effect.

Thus, in my final experimental chapter I returned to the food sharing context, exploring the flexibility of male Eurasian jays' sensitivity to their partners' desires. I also additionally conducted a comparative experiment with adult humans to address the question of a similarity between jays' and humans' desire sensitivity. In **Chapter 6,** I investigated how male Eurasian jays alter their food sharing patterns when their partner chooses between two identical food types (non-informative choice) or between two different food types (informative choice). Eurasian jays altered their food sharing pattern only when the choices were informative with regards to their partner's preference. In the human experiment I used a reaction times measure to determine whether adults' implicit processing of an agent's preferences might interfere with their ability to respond to a cue. I did not find evidence for an implicit response to the informativeness of others' choices; instead my results seemed to be more in line with a low-level difference in novelty rather than the attribution of mental states.

Finally, in **Chapter 7** I discuss my experimental findings with regard to the questions identified in this introductory chapter. I consider the limited evidence that human adults may have an implicit sensitivity to the preferences of others and whether this might stem from a consistent minimal system. I also discuss how my investigation of Eurasian jays' sensitivity to others' desires may cast doubt on our ability to assess their social cognitive abilities outside of the food sharing context. I conclude by assessing whether there may be evidence for a shared minimal theory of mind in Eurasian jays and human adults, considering this question in light of recent failed replications and criticisms of the underlying evidence for the minimal system.

2. Chapter 2 - Spontaneous responses to others' specific satiety in human adults

2.1 Introduction

As discussed in Chapter 1, traditional false belief tasks involve directly questioning children about behavioural outcomes and are passed around the age of 4 (Wellman & Liu, 2004). However, recent research with human infants has been strongly focused on spontaneous responses. These contrast with traditional false belief tasks in that they measure unprompted automatic responses to a scenario (Baillargeon et al., 2010). Tracking of eye movements has shown spontaneous looking responses to false beliefs in children as young as 10 months (Kovács et al., 2010; Surian et al., 2007); while other spontaneous measures, such as motor cortex activation, have also shown indications of belief processing in the first year of life (Southgate & Vernetti, 2014).

Research into spontaneous responses has sparked a new wave of investigation into the mechanisms underlying theory of mind, with two main interpretations emerging (Low & Perner, 2012). Early mind-reading accounts argue that spontaneous responses are indicative of a fully representative theory of mind (Baillargeon et al., 2010; Carruthers, 2013). In contrast, conceptual change accounts argue that there is a qualitative difference between the social cognitive abilities possessed by infants and adults (Apperly & Butterfill, 2009; Csibra & Gergely, 1998; Nichols & Stich, 2003; Ruffman et al., 2012).

One such conceptual change account is the two systems model described by Apperly and Butterfill (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013). This model suggests that spontaneous responses may result from a minimal theory of mind system that is separate to the system that generates explicit responses. This minimal system achieves cognitive efficiency by representing belief-like registrations as opposed to fully representing beliefs. Apperly and

Butterfill suggest that there may be some similarities between human minimal theory of mind and the mechanisms underlying the social cognitive abilities of non-human animals.

2.2 Comparing sensitivity to specific satiety in human adults and Eurasian jays

To my knowledge, no existing research has investigated spontaneous eye movement responses to desires in human adults. In contrast, research into theory of mind abilities in non-human animals has included the investigation of Eurasian jays' sensitivity to others' desires (Ostojić et al., 2013). Eurasian jays' co-operative social cognitive abilities have been investigated using their natural food sharing courtship behaviour, in which males place food directly into the beak of their female partner (Ostojić, Cheke, Shaw, Legg, & Clayton, 2016; Ostojić, Legg, et al., 2016; Ostojić, Legg, Shaw, et al., 2014; Ostojić et al., 2013). In these studies, the female's desire was manipulated via specific satiety, a sensory phenomenon whereby continued consumption of a single food causes a temporary decline in the desirability of that food (Balleine & Dickinson, 1998; Havermans, Janssen, Giesen, Roefs, & Jansen, 2009). It was found that the males shifted their pattern of food sharing in line with the females' desires, such that they shared a greater proportion of waxworms with their partner when she was pre-fed mealworms than when she was pre-fed waxworms (Ostojić et al., 2013).

In order to investigate the similarities between human minimal theory of mind and desire attribution in non-human animals, here I have adapted the specific satiety paradigm used with Eurasian jays to test the spontaneous responses of adult humans to others' desires.

2.3 Experiment 1

In the Eurasian jay paradigm, the females were pre-fed either wax moth larvae (*Galleria mellonella)*, henceforth referred to as waxworms, or mealworm beetle larvae (*Tenebrio molitor*), henceforth referred to as mealworms. This was then followed by a food sharing test phase during which the male had the opportunity to share up to 20 waxworms or mealworms with the female. To modify this for adult humans, participants watched videos of a protagonist eating either apples or crackers, followed by a video clip in which the protagonist was offered a choice between apples and crackers. This experimental design was influenced by an earlier study conducted with children, which involved participants choosing which type of food (apples or crackers) to give to a specifically sated protagonist (Legg, 2014a).

If the participants were able to spontaneously process the protagonist's choices based on their satiety they should anticipate the protagonist choosing the non-sated food. If directed anticipatory looking is found this may indicate a similar sensitivity to specific satiety shared

between adult humans and Eurasian jays. In contrast, if no directed anticipatory looking is found this would not provide support for a similarity between a human minimal system and animals' theory of mind abilities.

2.3.1 Participants

I chose a sample size of 25 participants as this was slightly larger than those used in previous studies that have identified anticipatory looking responses in adults (19-20) (Low & Watts, 2013; Schuwerk, Vuori, & Sodian, 2015; Senju, Southgate, White, & Frith, 2009). To achieve this I recruited 35 adults, 10 of these were excluded due to software error (1), experimenter error (i.e. sound was off, 2) or a lack of eye movements (i.e. looking straight ahead throughout the test, 7). This left 25 participants, 20 female, mean age 25 years 6 months, age range 20 years 8 months to 39 years 8 months. Participants were recruited through online and poster advertisements and paid £4 for taking part in the experiment. After reading an information sheet, all participants gave written consent. This experiment was approved by the Department of Psychology Ethics Committee (PEC) at the University of Cambridge.

2.3.2 Methods

Participants were tested in the Department of Psychology at the University of Cambridge, between July and November 2016. Participants were seated in front of a 27-inch screen allowing unconstrained viewing at a distance of approximately 1m. To allow gaze direction to be determined participants were recorded during the trial using a Logitech webcam and Camtasia software. All participants were presented with two conditions (an apples condition and a crackers condition), and their responses to the two conditions were compared in a within-subject design. The order of the two conditions was counterbalanced across participants.

Each condition involved a video being presented to the participant. The videos comprised two sections: an eating clip (Figure 2.1a) and a choice clip (Figure 2.1b). The eating clip showed a protagonist eating from a bowl of food: apple slices in the apples condition and Ritz crackers in the crackers condition. The choice clip was identical between the two conditions: the protagonist was presented with two bowls, one containing apples and one containing crackers. The sex of the protagonist in the videos was matched to the sex of the participant (Griskevicius et al., 2007). To prevent carryover effects, the videos were captioned 'Day 1' before the first video and 'Day 2' before the second. This was intended to indicate that the videos were depicting the protagonist's actions on two separate days, and thus the protagonist's satiation state in the second video should be independent of the first video.

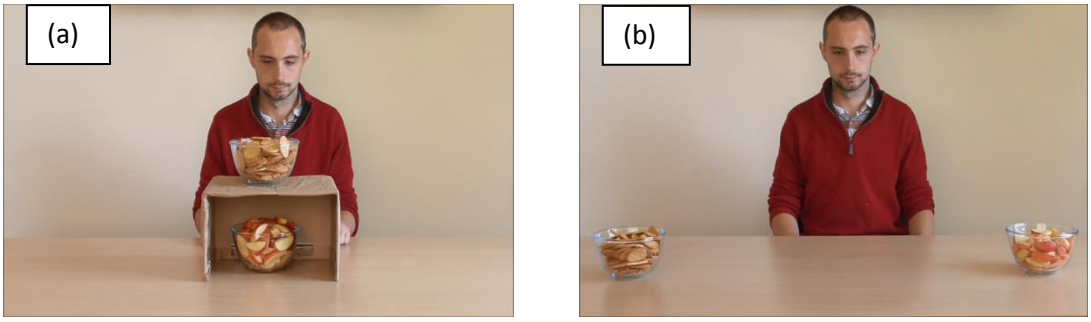

Figure 2.1: Stills taken from the videos showing the male protagonist. Throughout the four minute eating clip (a) the protagonist was shown eating either crackers or apples; the other food-type was hidden below a box so it was visible to the participant but not the protagonist. In the choice clip (b) the protagonist was presented with separate bowls of the two different food-types.

Each eating clip was four minutes in length. A pre-test conducted by Legg before his study with children found this to be long enough for participants to identify that the protagonist was sated (Legg, 2014a). Food items were eaten at a constant rate of one item every 15 seconds, and protagonists maintained a neutral expression throughout. The apples were cut into slices of a similar size to the Ritz crackers, and both food items took approximately the same amount of time to eat. A beep occurred each time the protagonist reached towards the bowl, so that the participant would learn to expect a reaching movement when the beep was played. This is analogous to the light and tone used in spontaneous false belief studies to cue participants to anticipate a reaching movement (Senju et al., 2011, 2009; Southgate, Senju, & Csibra, 2007).

To prevent participants from visually habituating to the presence of one food during the eating clip, the other food was placed under a box with an open side facing the participant. The bowl was therefore visible to the participant but not visible to the protagonist (Figure 2.1a). This should ensure that any difference in looking preference towards the non-sated food during the choice clip would not be explained by perceptual novelty. The other bowl of food was not visible to the protagonist so it should not appear to the participant as if the protagonist is choosing one food over the other.

After watching the 4-minute eating clip, participants were immediately presented with the choice clip. The protagonist entered the room and sat facing the participant with two bowls - one containing apples and one containing crackers - on a table in front of them (Figure 2.1b). In order to increase the chances of an anticipatory looking response, the beep used in the eating clip was repeated. The protagonist then leant forward in a 1 second clip and the video was paused just before the protagonist would have reached for the bowl, displaying the final image statically for 5 seconds (Senju et al., 2009). The two bowls were placed as far apart as possible so

that eye movements would be easier to code, and the positions of the two foods were counterbalanced across participants.

2.3.3 Recording of responses

The participants' looking responses were measured by coding the video recordings after the trials had concluded. For each participant two measures of anticipatory looking were assessed from the same recordings: (i) the initial direction in which the participant looked, (ii) the total time spent looking towards each bowl of food (Low & Watts, 2013; Schuwerk et al., 2015; Senju et al., 2009). Frame-by-frame video was exported into VLC player and looking directions were recorded through experimenter observation. Twenty percent of videos (5 out of 25) were coded by a second experimenter who was blind to condition. Inter-rater reliability was high, with 96% agreement over the occurrence of looks. The direction of first looks provided 100% agreement, and the difference between experimenters' measurements of looking time to each food-type was just 8.1% on average (0.050 ± 0.048 seconds, mean ± standard deviation).

At the end of the experiment participants' explicit responses were recorded by asking them to mark their expectation for the protagonist's choice along a line from apples to crackers (Figure 2.2), giving a quantitative representation of their prediction. Participants were given two lines, one for each video. The position of the choices (left and right) was randomised and counterbalanced between participants. The continuous response mirrored the continuous nature of the looking time measure and allowed participants to give an intermediate answer if they were not confident of which food the protagonist would choose. The explicit questions were given to the participant after the presentation of both of the videos to prevent their explicit answers from influencing their eye movements.

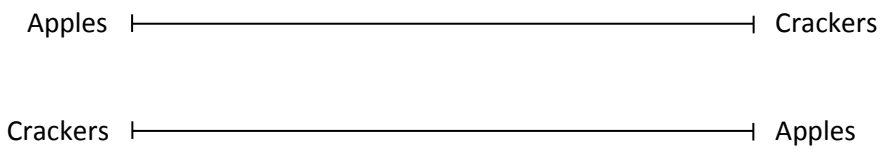

Figure 2.2: The lines as they appeared to the participant when giving their explicit answer. Participants were given lines in one of the two orders, apples to crackers or crackers to apples. Each participant was presented with two lines, one for 'Video 1' and one for 'Video 2'. The explicit question for both conditions was given to participants after they had watched the second video. Participants were asked to mark along the lines how likely they though the protagonist was to choose either food, the closer to one end of the line the more confidence they showed that the protagonist would choose this food.

2.3.4 Analysis

To analyse the participants' understanding of the protagonist's desire, a within-subject comparison was made between each participant's anticipatory looking responses in the two conditions; apples and crackers. Previous studies looking at spontaneous responses to false belief tasks have typically compared looking responses to 'chance' (usually defined as 0.5), assuming that uninformed participants are equally likely to look to each location. However, this does not account for baseline variation in looking preferences across participants. By comparing across two trials it was possible to control for disparities in looking times across locations due to visual differences in colour, personal preference for either food, handedness, or other underlying side biases. This should also prevent differences in eye movements from being affected by salient visual features of the scenario as they should be constant across conditions (Kulke, von Duhn, Schneider, & Rakoczy, 2018). When the difference between the two conditions is measured this should extract only the variation resulting from anticipation of the protagonist's choice. This also reflects the analysis used in the Eurasian jays experiments, in which the male's food sharing was compared in a within-subject design between a pre-fed waxworms condition and a pre-fed mealworms condition in order to control for variations in the males' baseline preferences for sharing different foods (Ostojić et al., 2013).

The pattern of first looks was compared between the two conditions (apples and crackers) using a McNemar's test for consistency of responses. The proportion of time spent looking towards apples out of total time spent looking at apples and crackers[8] was compared between the two conditions using a Wilcoxon signed-rank test. If participants were able to spontaneously anticipate the protagonist's choice of the non-sated food, they should spend a higher proportion of time looking at apples in the crackers condition than the apples condition.

Participants' explicit answers were also analysed using a within subject design as they again may be influenced by the participant's own baseline preference for each food type. The answers were scored numerically by measuring along the line from apples to crackers, with an accuracy of 0.05cm. Their answers were compared between the two conditions using a Wilcoxon signed-rank test. If participants were able to explicitly predict the protagonist's choice of the non-sated food, they should mark closer to apples for the crackers condition than for the apples condition.

[8] The proportion of total looking time directed towards apples was derived using the formula $\frac{A}{A+C}$, where A is the time spent looking towards apples and C is the time spent looking towards crackers. The proportion of total looking time directed towards crackers is the inverse of this value and therefore does not alter the output of the analysis.

Standardised effect sizes were calculated for the looking time measure and the explicit question using Cohen's *d* (mean difference/standard deviation). These effect sizes may be useful for comparing relative effect size between experiments however they are unlikely to provide an accurate measure of the true effect size as the sample sizes are small and the assumptions are likely to be violated.

2.3.5 Results & Discussion

In response to the explicit question, participants indicated that the protagonist was more likely to choose apples after having eaten crackers than after having eaten apples (Wilcoxon signed-rank test, n = 25, V = 234, p = 0.0150, Cohen's d = 0.851) (see Appendix C for summary statistics). In contrast, participants' eye movements (spontaneous measure) showed no directed response to the protagonist's specific satiety. Participants showed no difference in looking times towards apples between conditions (Wilcoxon signed-rank test, n = 25, V = 121, p = 0.540, Cohen's d = 0.184, Figure 2.3), and were equally likely to direct their first look towards apples in both conditions (McNemar's test, n = 25, X^2 = 0.1, p = 0.752, Figure 2.3). These results suggest that participants did not spontaneously process specific satiety despite making explicit predictions that the protagonist would choose the non-sated food.

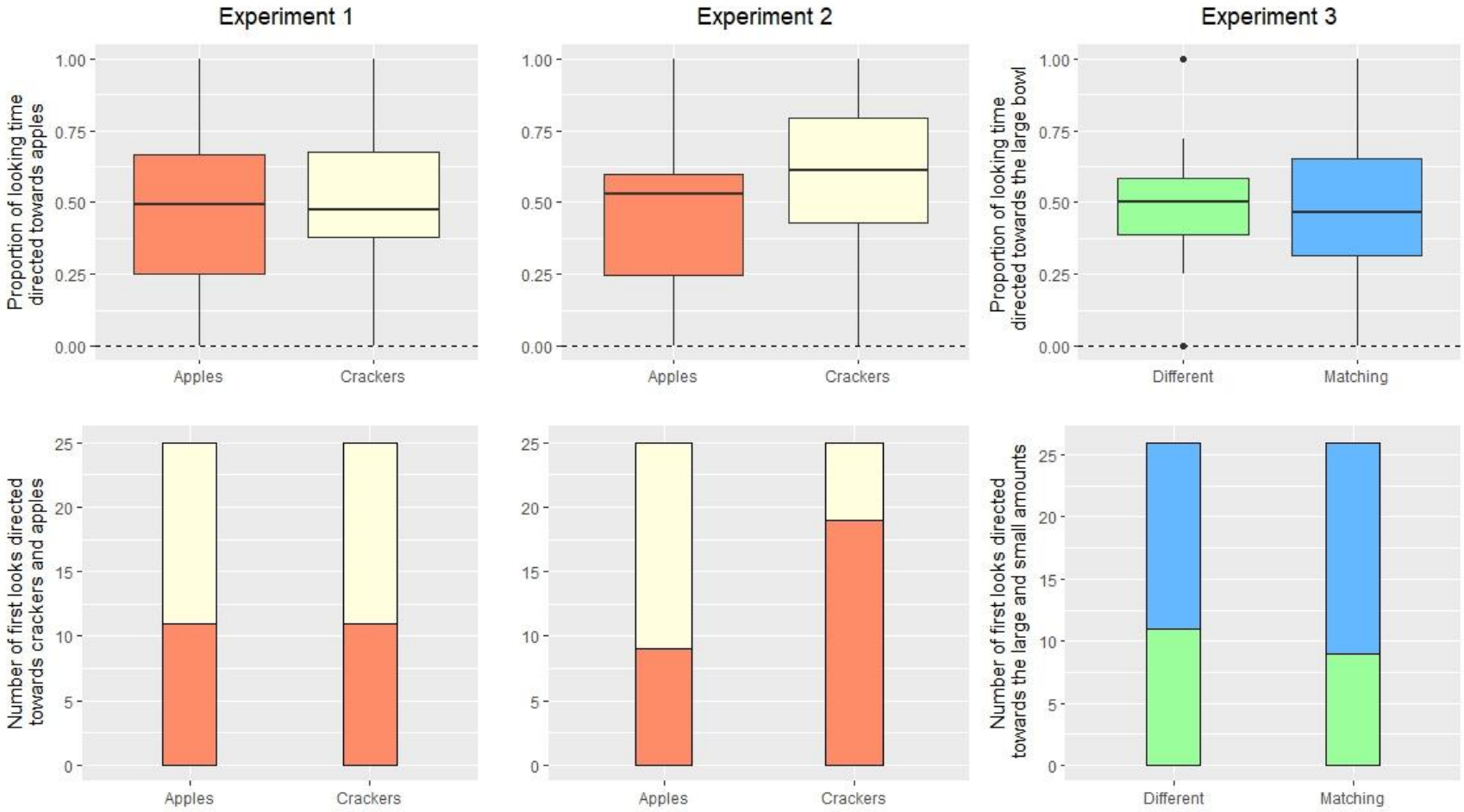

Figure 2.3: Spontaneous responses of participants to videos: The boxplots show looking time measures for the two conditions, apples eaten (red) and crackers eaten (yellow) for Experiments 1 and 2, different protagonist (green) and matching protagonist (blue) for Experiment 3. Boxes represent interquartile range and median, while whiskers represent the full range of the data. The barplots show the direction of first looks towards apples (red) and towards crackers (yellow) for Experiment 1 and 2, and towards the large amount (green) and the small amount of chocolate (blue) for Experiment 3.

2.4 Experiment 2

Subsequent to Experiment 1, I conducted a further experiment to investigate whether adults were able anticipate a protagonist's specific satiety when provided with a direct indication of the protagonist's desire. One key difference between the Eurasian jay study and the paradigm used in Experiment 1 is that I was able to exercise some control over behavioural indications of the protagonist's satiety. That is, by instructing the actor to keep a neutral expression throughout the eating clip and by timing the intervals so that food was consumed at a constant rate, I limited cues that could indicate satiety. This contrasts with the Eurasian jay study, in which the female was given free access to the pre-fed food for 15 minutes, making it possible that she demonstrated behavioural indications of her satiety during this time. As a result, the males' shift in sharing pattern may be due to behaviour reading of the female's satiety during the pre-feeding phase, as opposed to true desire attribution.

In Experiment 2 the paradigm was modified to include a direct behavioural indication of the protagonist's desire. A verbalisation, either "I'm full of apples" or "I'm full of crackers", was added at the end of the choice clip, similarly to a previous study conducted with children (Legg, 2014a). This additional cue should allow the participant to predict the protagonist's goal without needing to attribute a desire. It is highly likely that adults will have previously encountered scenarios in which other individuals (or they themselves) have said they are full of a food and subsequently chosen not to eat it. Thus, the verbalisation provides the potential to bypass desire attribution, allowing participants to predict the protagonist's choices through behaviour reading and previous experience.

2.4.1 Participants

Participants were 36 adults, 11 of these were excluded due to experimenter error (2), software error (4), participant not watching videos to completion (1), or a lack of eye movements (4). Of these 25, 15 were female and the mean age was 25 years 3 months, with an age range of 19 years 2 months to 39 years 8 months. Participants were recruited through online and poster advertisements and paid £4 for taking part in the experiment. After reading an information sheet, all participants gave written consent. This experiment was approved by the Department of Psychology Ethics Committee (PEC) at the University of Cambridge.

2.4.2 Methods & Analysis

Participants were tested in the Department of Psychology at the University of Cambridge, between July and November 2016. The method was the same as for Experiment 1 except that the desire of the protagonist was directly indicated through the addition of a verbalisation at the end of each of the videos: "I'm full of apples" or "I'm full of crackers", for the apples and crackers conditions respectively. The analyses were identical to Experiment 1. Twenty percent of videos (5 out of 25) were coded by a second experimenter who was blind to condition. The inter-rater reliability was very high, with 100% agreement for both the occurrence and direction of looks. The difference between experimenters' measurements of looking time to each food-type was just 5.0% (0.035 ± 0.049 seconds).

2.4.3 Results & Discussion

As in Experiment 1, participants explicitly answered that the protagonist was more likely to choose apples after eating crackers than after eating apples (Wilcoxon signed-rank test, n = 25, V = 279, p = 0.00103, Cohen's d = 1.14) (see Appendix C for summary statistics). Participants also did not direct a different proportion of their total looking time towards apples in the two conditions (Wilcoxon signed-rank test, n = 25, V = 105, p = 0.208, Cohen's d = 0.437, Figure 2.3). However, in contrast to Experiment 1, the direction of participants' first looks was dependent on what the protagonist had eaten. Participants were more likely to look first towards apples after the protagonist had eaten crackers than after the protagonist had eaten apples (McNemar's test, n = 25, X^2 = 5.79, p = 0.0162, Figure 2.3). These results suggest that the addition of a verbalisation of the protagonist's desire may have allowed participants to implicitly anticipate the protagonist's choice.

The lack of a significant difference between the conditions in the looking time measure may be a result of the relatively small sample size used in this experiment, as well as increased noise caused by the manual measurement of looking times as I did not have access to an eye tracker. However, it is also possible that the lack of consistency in the results between the spontaneous measures is due to a false positive finding for first looks.

2.5 Experiment 3

One possible explanation for the results seen in Experiment 1 and 2 is that when no specific verbalisation was made the participants spontaneously attributed general satiety to the protagonist towards both foods. The participants therefore wouldn't expect the protagonist to choose either of the two foods, leading to a lack of anticipatory looking. In contrast, the explicit

question may have prompted the participants that the protagonist would choose at least one of the two foods, thus leading them to answer based on specific satiety. In Experiment 2, the addition of the verbalisation may have prompted anticipatory looking as only one food type was named, making the specificity of the satiety more salient.

In Experiment 3, I therefore investigated the possibility that participants are able to spontaneously attribute general satiety towards a single food type. Participants watched videos of a protagonist eating a bowl of chocolate. This was followed by a video clip in which either the same protagonist or a differed protagonist was offered a choice between a large bowl of chocolate or small bowl of chocolate. If participants are able to spontaneously attribute general satiety they should anticipate that a sated protagonist would be more likely to choose a small amount of food over a large amount of food. In contrast, a non-sated protagonist should be more likely to choose a large amount of food.

2.5.1 Participants

Participants were 35 adults, 9 of these were excluded due to experimenter error (2), participant not watching videos to completion (1), or a lack of eye movements (6). This left 26 participants, 17 female, mean 23 years 3 months, age range 18 years 8 months to 41 years 0 months. The sample size was intended to be 25 as in the other two experiments; however fewer participants were excluded than anticipated. Participants were recruited through online and poster advertisements and paid £4 for taking part in the experiment. After reading an information sheet, all participants gave written consent. This experiment was approved by the Department of Psychology Ethics Committee (PEC) at the University of Cambridge.

2.5.2 Methods & Analysis

Participants were tested in the Department of Psychology at the University of Cambridge, between November 2017 and January 2018. Due to software errors in Experiment 1 and 2 caused by the high processing requirements of the Camtasia software, in Experiment 3 participants were recorded with a GoPro camera in addition to the Logitech webcam. Videos were presented to the participants in the same manner as in Experiments 1 and 2. Each video consisted of two sections: an eating clip in which a protagonist ate a bowl of chocolate, and a choice clip in which a protagonist was seated between two bowls, one containing a small amount and one containing a large amount of chocolate. The sex of the protagonist was matched to the sex of the participant.

Each eating clip was four minutes in length, with chocolate squares eaten at a constant rate of one every 12 seconds and protagonists maintaining a neutral expression throughout (Figure 2.4a). The bowl contained 20 squares of chocolate so that the bowl was empty at the end of the 4-minute period. This was important to increase the likelihood that participants would expect the protagonist to choose more chocolate during the choice clip. If chocolate remained in the bowl at the end of the eating period this may indicate a rejection of the food, and thus it might appear unlikely to the participant that the protagonist would choose to subsequently eat more chocolate. As in Experiments 1 and 2, a beep occurred each time the protagonist reached towards the bowl, so that the participant would learn to expect a reaching movement when the beep was played.

In the choice clip, the protagonist entered the room and sat facing the participant with two bowls - one containing 20 squares and one containing 5 squares of chocolate - on a table in front of them (Figure 2.4b). In order to cue the participants' anticipatory looking, the beep used in the eating clip was repeated. The protagonist then leant forward in a 1 second clip and the video was paused for 5 seconds just before the protagonist would have reached for the bowl. The two bowls were placed as far apart as possible to increase accuracy of coding, and the contents of the two bowls were counterbalanced across participants.

A pre-test was conducted on the online platform Prolific using stills from the choice section of the video to determine which bowls participants would expect an individual to choose to eat when sated or not sated. Participants from the pre-test were not used in the experiment. As expected, participants answered that they would choose to eat the large amount themselves (Wilcoxon signed-rank test, two-tailed, $n = 36$, $p < 0.01$) and that they would expect someone else to choose the large amount ($p < 0.01$), but that they would expect someone who had already eaten a large amount of chocolate to choose the smaller amount ($p < 0.05$).

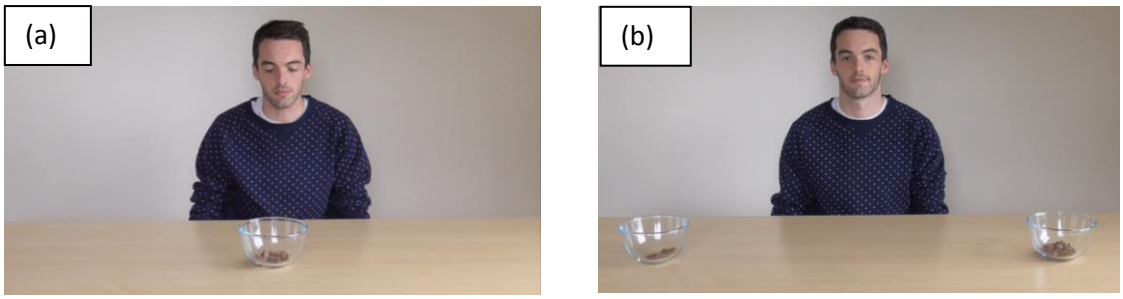

Figure 2.4: Stills taken from the videos showing the male matching protagonist. Throughout the four minute eating clip (a) the protagonist was shown eating chocolate pieces until the bowl was empty. In the choice clip (b) the protagonist was presented with separate bowls, one containing a large amount of chocolate, and the other a small amount of chocolate.

In the main experiment, participants' responses were compared in a within-subject design, between two different conditions: a matching condition and a different condition. A within-subject design was chosen here for the same reasons as outlined above in Experiment 1, namely to eliminate looking differences due to factors other than the difference in satiety (e.g. visual differences between the small and large bowls of chocolate). The two conditions differed only in the identity of the protagonist.[9] In the matching condition participants watched an eating clip featuring protagonist A, followed by a choice clip featuring protagonist A. In the different condition participants watched an eating clip featuring protagonist B, followed by a choice clip featuring protagonist A. The order of presentation of the conditions was counterbalanced across participants. The identities of protagonist A and protagonist B were kept consistent across participants.

It was predicted that when participants saw a protagonist who had already eaten (matching condition) they would expect them to choose the smaller amount, but when they saw a novel protagonist being given the choice (different condition) they would expect them to choose the larger amount. Comparing between these two conditions, with identical choice clips, should

[9] Studies of implicit understanding goal-directed action in infants have shown that they are sensitive to the identity of the protagonist (see Chapter 1). For example if a protagonist is shown reaching for object X over object Y in familiarisation trials, infants will show a violation of expectation response if they are then shown the protagonist reaching for object Y (Woodward, 1998). However, they do not show such a response if the protagonist has changed to a different individual, presumably because they have no reason to expect the novel protagonist to have the same preference (Buresh & Woodward, 2007; Henderson & Woodward, 2012).

allow extraction of only those eye movement differences that are related to the identity/satiety of the protagonist.

Participants' responses were recorded in the same way as in Experiments 1 and 2. Spontaneous responses were coded from the video recordings after their trials had concluded. Explicit responses were recorded by asking participants to mark their expectation for the protagonist's choice in each video along a line from small to large. The answers were scored numerically by measuring along the line with an accuracy of 0.05cm. The position of the choices was counterbalanced between participants. Analyses were conducted in the same way as in Experiments 1 and 2, comparing responses between the matching and different conditions in a within-subject design. Nineteen percent of videos (5 out of 26) were coded by a second experimenter who was blind to condition. The inter-rater reliability was very high, with 100% agreement over the occurrence and direction of looks. The difference between experimenters' measurements of looking time to each food-type was just 6.7% on average (0.046 ± 0.102 seconds).

2.5.3 Results & Discussion

Participants indicated that the protagonist was more likely to choose the small amount when the protagonist was sated (matching condition) than when the protagonist was not sated (different condition) (Wilcoxon signed-rank test, n = 26, V = 97.5, p = 0.0470, Cohen's d = 0.605) (see Appendix C for summary statistics). Across the two conditions, participants showed no difference in the proportion of total looking time directed towards the small amount (Wilcoxon signed-rank test, n = 26, V = 176, p = 0.692, Cohen's d = 0.0203, Figure 2.3), and no difference in first look direction (McNemar's test, n = 26, X^2 = 0.1, p = 0.752, Figure 2.3). Thus, the participants did not appear to anticipate a difference in behaviour between a sated and a non-sated protagonist, despite showing an explicit response in line with the protagonists' satiety.

However, there are some issues with this experimental design that make these findings more difficult to interpret. Specifically, it may not have been clear in the choice clip that the protagonist was making a choice between the small amount and the large amount of chocolate. Both bowls contained chocolate so if the protagonist was reaching for a single piece of chocolate they could reach for either bowl. This contrasts with Experiments 1 and 2, where the participant was presented with one bowl of apples and one bowl of crackers, so if the protagonist was reaching for a single piece of food they would have to reach towards the appropriate bowl.

Experiment 3 therefore only tests the participant's understanding of the protagonist's satiety if the participant assumes the protagonist was reaching for the bowl itself, or reaching for the

chocolate with the intention of eating the entire contents of the bowl. As such, the lack of a directed anticipatory looking response could have been due to ambiguity in the videos, rather than a failure to attribute satiety. In contrast, in the explicit question and the pre-test, participants were asked which bowl of chocolate the protagonist would choose (large or small) making these questions less ambiguous.

2.6 General Discussion

Across three experiments I tested whether human adults show spontaneous anticipatory looking in line with others' satiety. In all three experiments participants were able to make explicit predictions based on a protagonist's satiety. However, their spontaneous responses differed across experiments. In Experiment 2, which included a direct verbalisation of the protagonist's satiety "I'm full of apples/crackers", participants seemed to spontaneously anticipate the protagonist's choice in one of the two measures. In contrast, in Experiments 1 and 3, where participants had no overt cues about the protagonist's desire, anticipatory looking did not appear to be influenced by the protagonist's satiety. It is unlikely that the lack of an anticipatory response in Experiments 1 and 3 was due to a failure of the choice clip to trigger anticipatory looking, given that a directed response was found in Experiment 2. In addition, no directed anticipatory response was found in Experiment 3, providing no evidence to suggest that the results of Experiment 1 could be explained by attribution of general satiety. Although this finding must be tentative given the possible ambiguity of the videos used in Experiment 3.

Together these three experiments do not provide concrete evidence for a spontaneous attribution of desires to others' in adult humans. Participants only appeared to be able to spontaneously anticipate the protagonist's choice when their desire was indicated by direct vocalisation, thus allowing for an experience-based behaviour-reading explanation. Furthermore, Apperly and Butterfill have suggested that their minimal system may resemble the strategy used by non-human animals to solve tasks that appear to involve tracking others mental states (Apperly & Butterfill, 2009). The lack of anticipatory response from human adults in Experiment 1 may therefore be surprising given the success of Eurasian jays in a specific satiety food sharing paradigm.

However, in this paradigm the behavioural cues available to the participants were very different to those available to the Eurasian jays in the original paradigm (Ostojić et al., 2013). During the pre-feeding phase of the jay experiments, the behaviour of the female may provide cues to the male regarding her satiety. In Ostojić et al.'s experiment the male was found to only shift his pattern of food sharing when he was able to watch the female eating (seen condition), and not

when his visual access to her was blocked during pre-feeding (unseen condition). This control showed that the female's behaviour during food sharing alone was not sufficient to change the male's sharing pattern. However, it does not rule out an effect of differences in the female's behaviour during pre-feeding. It is possible that as the female becomes sated on the pre-fed food her eating behaviour changes in some way (e.g. slower rate of eating), providing an observable indicator of her internal state. Hence, the male jays may be able to utilise a behaviour reading strategy in order to determine which foods to share, rather than requiring true desire attribution.

In contrast, in the human paradigm, I tried to minimise the cues available to the participant by keeping the expression of the protagonist neutral and the rate of eating constant. As behavioural rule accounts can be constructed *post hoc* it is always possible to explain the outcome of any paradigm with a behavioural rule (Lurz, 2009). However, if obvious behavioural cues are minimised the effect size of any anticipatory response may be reduced, making it much more difficult to identify an effect in Experiment 1 where behavioural cues were limited.

Behavioural cues have also been shown to be important to young children when reasoning about others' satiety. In a previous study, children were asked to choose between apples and crackers to share with a protagonist who had been sated on one of the foods (Legg, 2014a). Children under six years old were unable to respond correctly when they had to infer the protagonist's satiety from the food they had eaten, but were able to respond correctly when the protagonist provided a verbal indication of their desire. Overall, it is possible that the adult humans and Eurasian jays do share similar theory of mind abilities in the context of specific satiety, both sensitive to the behavioural cues of other's satiety in their food sharing and implicit responses respectively. However, based on these experiments alone it is not possible to disambiguate a minimal theory of mind explanation from alternative behaviour-reading explanations.

Furthermore, there has been recent criticism of the use of looking measures, such as violation of expectation and anticipatory looking, to assess implicit theory of mind. Firstly, since the experiments in this thesis were conducted (2016/17), a series of studies have been published that indicate that evidence for spontaneous responses to false beliefs is less strong than previously believed (Kulke & Rakoczy, 2018). A range of replication studies have failed to find evidence for the spontaneous responses previously identified in young infants (Burnside, Ruel, Azar, & Poulin-Dubois, 2017; Dörrenberg, Rakoczy, & Liszkowski, 2018; Kulke, Reiß, Krist, & Rakoczy, 2017; Kulke, von Duhn, et al., 2018). As such, questions have been raised about the

validity of these measures, and whether they are appropriate methods for assessing implicit theory of mind (see Chapter 7 for further discussion).

Secondly, one of the most significant supports for Apperly and Butterfill's two system account has been the identification of certain signature limits that are theoretically predicted by their minimal model. The primary example of this is the inability of the minimal system to recognise *how* an individual perceives an object, i.e. second order perspective taking (Low & Watts, 2013) (see Chapter 1). However, recent criticisms have been made of the original study in which this signature limit was tested. Low and Watts's study compared infants' spontaneous responses in a traditional object-location test and a novel object-identity test probing second order perspective taking. Kulke et al. were initially able to replicate the results of the object-location test, but they suggested that this outcome may be due to salient visual cues during the videos (Kulke, von Duhn, et al., 2018).

In a further replication that included control measures, they found that participants' looking responses were significantly confounded with the direction of the protagonist's movements. Thus, the original differences in spontaneous responses between the object-location and object-identity tests could potentially be explained by factors unrelated to their perspective taking requirements. Thus, the evidence for signature limits that align with Apperly and Butterfill's two systems account may be less robust than previously believed. In addition, this also coincides with other criticisms of looking measures, which suggest that the differences found in these studies may be attributed to lower level explanations or artefacts of the experimental design rather than implicit theory of mind (Heyes, 2014a, 2014b; Perner & Ruffman, 2005; J. Phillips et al., 2015).

Finally, a major experimental support for Apperly and Butterfill's suggestion of the similarity between minimal theory of mind and animal mental state attribution is provided by a recent demonstration of anticipatory looking behaviour in apes (Krupenye et al., 2016). However, some criticisms have been made of the interpretation of these experiments. Heyes has argued that the results could be explained by 'submentalising', in which characteristics of the stimuli cue attention in such a way that spontaneous responses are generated without any requirement for mindreading (Heyes, 2017). In order to test for such an explanation Heyes suggested a non-social control condition, in which actors are replaced by inanimate stimuli. In response to this, Krupenye et al. conducted a further study in which they used non-social stimuli in the videos shown to the apes. In this study they found no evidence for directed anticipatory looking towards the box that a green shape should 'believe' a stone to be in. From this they concluded

that there was no evidence that submentalising was responsible for their results in their original experiment.

However, this conclusion may be too strong given that the non-social stimuli actually did show a trend for more looks in the direction that aligned with the beliefs, although this trend was not significant. This is particularly important here as the study did not directly test for a difference between the original anticipatory looking paradigm and the modified non-social paradigm. The results from the non-social condition (14/22 first looks to the correct location, 63.6%) were very similar to the original study (20/30, 66.7%), and it is incorrect to infer that a significant result will always differ from a non-significant result (Gelman & Stern, 2006). Furthermore, the experiment the authors chose to reproduce with non-social stimuli was Experiment 1 from the original study (Krupenye et al., 2016), in which the apes' probability of looking first towards the correct location did not differ from chance, rather than Experiment 2 in which this did differ from chance. This was an unusual choice by the authors and may have skewed the results in favour of their theory of mind explanation rather than Heyes's submentalising hypothesis.

In summary, although the experiments I have conducted in Chapter 1 don't rule out the possibility of a similarity between the theory of mind abilities of Eurasian jays and human adults it also does not provide unambiguous evidence that these responses are governed by a minimal theory of mind system. It is equally valid to consider that the implicit responses of human adults and the food sharing responses of Eurasian jays may be governed by different mechanisms. Furthermore, there are a growing number of concerns regarding the use of looking measures for assessing implicit theory of mind across human infants, adults and non-human animals. This further complicates the interpretation of these findings as it raises questions about the validity of such measures and whether anticipatory looking can, in principle, provide evidence for implicit theory of mind. In Chapter 2 I therefore explore further the possibility of implicit responses to others' desires and preferences in human adults using an alternative paradigm that avoids looking measures.

3. Chapter 3 - Implicit transitive inference of preferences in human adults

3.1 Introduction

It is difficult to investigate the mechanisms underlying an understanding of others' desires as there are often valid alternative explanations. If an individual takes an object this may indicate their desire for that object, but the individual is also now associated with that object. Thus any anticipation that the individual would subsequently choose the object could be based on this simple statistical association rather than the attribution of a desire or preference (Butterfill & Apperly, 2016; Ruffman et al., 2012) (see also Chapter 1). In Chapter 2, I addressed this associative issue using a specific satiety paradigm, in which prolonged consumption of a food leads to a decreased desire for that food. The food most associated with the protagonist is therefore the food that is *least* desired by the protagonist, and thus a simple associative explanation does not align with a desire-based explanation.

Here, in Chapter 3, I employed a transitive inference paradigm as an alternative solution to this association problem. Transitive inference is an ability, shared by humans and some non-human animals, that allows the relation between elements to be inferred from other relations without the need for explicit comparison. For example, if A is larger than B and B is larger than C, it can therefore be inferred that A is larger than C without comparing A and C with one another directly. Transitive inference experiments involve training participants on a series of pair-wise relations between stimuli in such a way that the pairs can be integrated into a continuous hierarchy. For example, the pairs A>B, B>C, C>D, D>E, with the symbol ">" representing an ordered relation such as larger than. These pairs imply the underlying 5-term hierarchy A>B>C>D>E. This can then be used to explore the attribution of preferences if each pair represents another individual's choice between the two stimuli; i.e. A>B indicates that A is chosen over B. When the terms B and D are compared, transitive inference should allow a participant to determine that B>D even though B and D have not been presented together previously. A simple associative explanation can be avoided as both B and D have been previously chosen by (and associated with) the individual.

3.1.1 Spontaneous transitive inference in infants

Recent studies employing spontaneous measures seem to provide evidence for a sensitivity to transitivity in young infants (Gazes, Hampton, & Lourenco, 2017; Mascaro & Csibra, 2014; Mou, Province, & Luo, 2014). These studies use a 'violation of expectation' paradigm, which relies on

the premise that infants will spend longer looking at an unexpected scenario than an expected scenario. Thus, if the infants are capable of constructing a transitive hierarchy from pairwise interactions (e.g. A>B>C), they should look longer towards a novel interaction that violates the hierarchy (e.g. A<C) than towards a novel interaction that is consistent with the hierarchy (e.g. A>C).

Crucially for my investigation of desire attribution in this thesis, one such study has indicated that infants may be sensitive to the transitivity of other's preferences. In Mou et al.'s study, 16-month-old infants were shown an agent choosing A over B, and then the same agent choosing B over C. The infants then looked longer at a video in which the agent chose C over A than one in which the agent chose A over C (Mou et al., 2014). The authors therefore concluded that infants transitively judged A to be preferred over C, based on the hierarchy A>B>C.

3.1.2 Implicit transitive inference in adults

There is also some existing evidence for an implicit response to transitive inference in adult humans. Adults are able to succeed on transitive inference tasks while being unable to explain their reasoning in terms of the underlying hierarchy (Greene, Spellman, Dusek, Eichenbaum, & Levy, 2001; Siemann & Delius, 1993). However, current research is limited to classic reinforcement-based transitive inference paradigms. Traditional experiments investigating transitive inference have used differential reinforcement of stimuli to create reinforcement-based hierarchies. Arbitrary stimuli are presented in pairs e.g. A+B- and B+C-, where "+" indicates that the stimuli is rewarded and "-" that the stimuli is unrewarded when chosen by the subject. From these pairs a reinforcement hierarchy A>B>C can be constructed and, if transitivity is inferred, the subject should choose A when it is paired with C.

Adult humans have been found to perform successfully on reinforcement-based transitive inference tasks without possessing explicit knowledge of the hierarchy. Using a 6-term hierarchy (A>B>C>D>E>F), Siemann and Delius found that 46.7% of the participants who performed high above chance were unable to report verbally how they had chosen their answers (Siemann & Delius, 1993). In a later experiment, Greene et al. trained participants on a 4-term hierarchy with half of the participants being told that the pairs could be ordered (informed group) and the other half told that they needed to learn the pairs by trial and error (uniformed group) (Greene et al., 2001). This training was followed by unreinforced test trials on the B vs D pair, and a post-experiment questionnaire to assess whether uninformed participants had gained awareness of the hierarchy. The performance of the informed group on the test was better than that of the uninformed group. However, within the uninformed group successful performance on the test

was unrelated to their answers to the questionnaire, suggesting that conscious awareness of the hierarchy was not required to successfully choose B over D. Thus, human adults appear to have the ability to succeed on reinforcement-based transitive inference paradigms without explicit logical reasoning.

3.1.3 Associative explanations for reinforcement-based transitive inference

Despite these previous findings, it is currently unclear what mechanisms might underlie adults' implicit transitive inference. There has been extensive discussion of the potential mechanisms for transitive inference in both humans and non-human animals; with cognition-based explanations competing with those based on associative processes (Allen, 2012; Vasconcelos, 2008). Cognitive explanations usually propose some sort of mental model developed during training, in which the full transitive hierarchy is represented, often spatially along a line (Zentall, 2001). This model then allows an individual to pass an inferential probe (e.g. B vs D) by judging which of the pair is closer to the end of the series.

In contrast, associative models argue that success on reinforcement-based paradigms may be possible without using transitive inference. The most basic associative explanations rely on the direct reinforcement history of the stimuli during training. For the simplest hierarchy A>B>C, participants are trained on the two pairs A>B and B>C, and then tested on the inferential probe A vs C. During training A is always rewarded (S+) and C is never rewarded (S-), providing a simple associative explanation for a participant choosing A over C. This is described as the 'end effect', and means that any pair that includes one of the end terms of the hierarchy can be explained by direct reinforcement history (von Fersen, Wynne, Delius, & Staddon, 1991). Thus, to avoid this effect transitive inference paradigms tend to employ hierarchies with at least 5 terms, as this is the shortest hierarchy that includes a transitive internal pair. Specifically, for the hierarchy A>B>C>D>E, the pair B vs D can be compared. The direct reinforcement history of B and D should be the same during training as each stimulus is reinforced 50% of time, B>C and D>E, and unreinforced 50% of the time, A>B and C>D.

However, it is not only the direct reinforcement history of the stimuli that can provide associative explanations, but also indirect reinforcement arising during training. The simplest model of indirect reinforcement is described by von Fersen's value transfer theory (VTT) (von Fersen et al., 1991). VTT suggests that the associative value a pigeon gives to a stimulus can also be affected by the value of the stimuli it is paired with. For example, as A is always reinforced it has a high value, and when A (S+) is paired with B (S-) some of this positive association transfers to B. As a result, the training gives B a higher value than D because B appears with higher value

stimuli (A and C) than D (C and E). This allows the pigeons' choice of B over D to be explained by associative processes rather than any understanding of the hierarchy. Subsequently, the principle of VTT was built upon to produce various associative models that accurately account for the pattern of results produced by reinforcement-based transitive inference paradigms (Siemann & Delius, 1998; Wynne, 1995).

Frank et al. suggest that previously identified implicit transitive inference in adults can be successfully predicted by associative models (Frank, Rudy, Levy, & O'Reilly, 2005). In this experiment, the authors used a 6-term hierarchy (A>B>C>D>E>F), and assessed participants on the internal pairs C vs E, B vs D and D vs E. Participants who were unaware of the hierarchy reliably chose B over E, but were at chance performance for C vs E and B vs D; a similar result to that seen previously in rats (Van Elzakker, O'Reilly, & Rudy, 2003). The authors argue that participants' performance on the test pairs seemed to be related to differences in associative strength, as predicted by the authors' model (Frank, Rudy, & O'Reilly, 2003); concluding that implicit success on reinforcement-based paradigms may be accounted for by relatively simple associative processes.

3.1.4 Implicit transitive inference outside of reinforcement-based paradigms

Despite the explanatory power of associative models, they focus only on reinforcement-based paradigms, and criticisms have been made regarding their generalisability. One such criticism is that reinforcement may not be an inherently transitive relation (Lazareva, 2012; Markovits & Dumas, 1992). The premises "A is reinforced, B is not", "B is reinforced, C is not" do not necessarily require that A is reinforced when presented together with C. This contrasts with inherently ordered relations such as "A is larger than B", "B is larger than C" as they imply positions along a linear scale and A must therefore be larger than C. As such, it is possible that there is a difference between the cognition required to succeed on reinforcement-based paradigms and that utilised in other forms of transitive inference.

Both of the infant studies described above (Gazes et al., 2017; Mou et al., 2014) conducted control experiments that failed to find associative explanations for the infants' responses. In contrast to the adult studies, these infant experiments involved dominance and preference hierarchies which appear to have inherent transitivity, i.e. if A is dominant over B, and B is dominant over C, it is then expected that A should be dominant over C. Therefore, following from these findings in infants, I investigated human adults' implicit transitive inference responses to inherently orderable preference-based hierarchies.

Although there is evidence suggesting that adults' implicit responses in reinforcement-based paradigms rely on associative mechanisms, this may not be the case for a preference-based hierarchy. If adult humans show an implicit response that, similarly to the infant studies, does not appear to be explained by associative mechanisms, this may provide evidence for a shared non-associative sensitivity to the transitivity of preferences.

3.2 Validation of Procedure

In order to assess adults' implicit transitive inference of a preference-based hierarchy I needed to first determine a valid implicit measure. In the infant experiments described above (Gazes et al., 2017; Mou et al., 2014), the infants' spontaneous responses were measured using a violation of expectation paradigm. However, given the current questioning of the reliability of eye movement based paradigms (see Chapter 2), including violation of expectation specifically (Heyes, 2014a); I chose to use a different method of assessing implicit understanding in this study.

In an experiment investigating young children's implicit understanding of false belief, Garnham and Perner used 'intentional action' as a measure of implicit knowledge (Garnham & Perner, 2001). Their paradigm resembled a traditional Sally-Anne task (see Chapter 1), except that the two boxes in which the object could be hidden were each placed at the bottom of a slide. In two separate sessions the children were either asked the explicit question "Where is Alan going to look?" or asked to move a mat to catch Alan at the bottom of the slide. They expected that in the "mat-moving" condition the children should be able to choose the slide that Alan was most likely to come down, without requiring explicit reasoning about his false belief.

The authors argue that their measure accesses implicit knowledge of false belief as the children are not being asked to consider the agent's beliefs directly, but instead performing an action to give their answer (moving the mat). Children's performance was higher on the mat-moving response than in their answers to the explicit question, and this performance was comparable to spontaneous looking responses within the same experiment. Thus, some active responses, such as mat-moving in the above study, may be governed by implicit knowledge, distinct from the explicit knowledge accessed by direct questioning.[10] I therefore chose to employ such an intentional action measure in this study.

[10] Repacholi and Gopnik's study, which tested children's understanding of diverse desires also used an intentional action-type measurement, with children being prompted to provide food to an experimenter rather than being asked a verbal question about the experimenter's desires (Repacholi & Gopnik, 1997) (see Chapter 1 for a full description of this study). Other studies have also used helping-related intentional

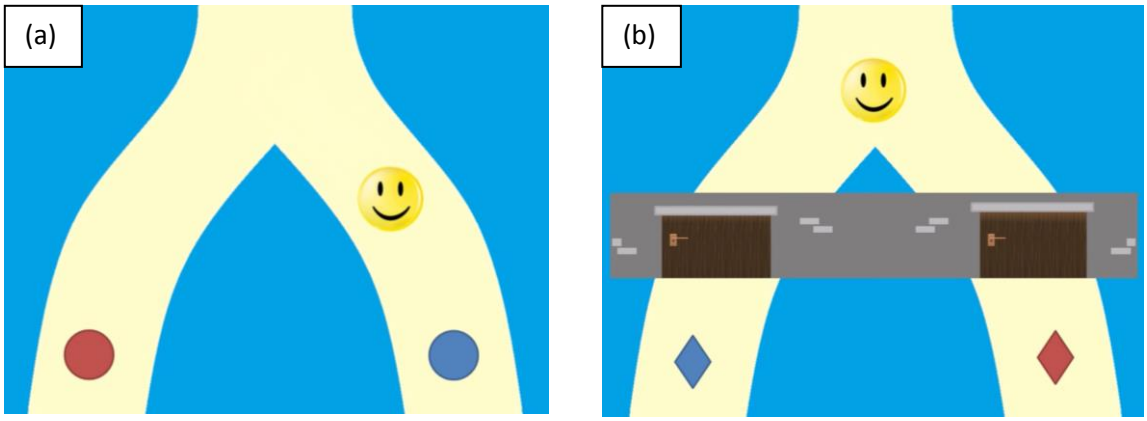

Figure 3.1 Stills taken from the videos shown to the participants. In the training videos (a) the agent was shown travelling down one fork of a path to reach an object. In the test videos (b) the agent was shown travelling down a path when a wall appeared blocking the way. Participants were then asked to indicate which door they would like to open for the agent.

However, prior to conducting the transitive inference element of this study (Experiments 1 and 2 below) I needed to validate the methodology of this paradigm, ensuring that my intentional action test question successfully prompted the participants to indicate the agents' preference. In this paradigm I showed participants videos (created using Microsoft PowerPoint) of an animated agent making a series of choices at forked paths, where each path led to a different object (see Figure 3.1a). In each video the choice of which path the agent travels down should indicate which of the two objects the agent prefers. At each fork the agent made a choice between objects that were identical except that one was red and one was blue. The agent consistently chose one colour over the other, which should be an easy rule for the adult participants to extract and thus a high level of success was expected (Low & Watts, 2013). These training videos were followed by a test video in which the agent's path was blocked by a wall with a door leading to each path (see Figure 3.1b). Participants were then asked which door they would like to open for the agent without specifically referencing the agent's preferences.

---

actions rather than verbal questioning to investigate understanding of false belief in young children (D. Buttelmann, Carpenter, & Tomasello, 2009; D. Buttelmann, Over, Carpenter, & Tomasello, 2014; F. Buttelmann, Suhrke, & Buttelmann, 2015; Southgate, Chevallier, & Csibra, 2010). However, there have been some criticisms of these paradigms, and failures to replicate their results (J. W. P. Allen, 2015; Dörrenberg et al., 2018; Poulin-Dubois et al., 2018).

3.2.1 Methods & Analysis

The videos were presented to the participants through an online survey hosted by Qualtrics. Each training video showed an animated agent (yellow smiley face) travelling down a path that split in two at a fork (Figure 3.1a). Each arm of the fork had an object of the same shape at the end of it, one of which was red and the other blue. Each participant was shown four such videos each with different shaped objects at the ends of the paths (triangle, star, square and circle). The videos were presented in a randomised order and in each video the location of the preferred shape (left or right) was also randomised. The colour preferred by the agent (red or blue) was counterbalanced across participants.

In the test video, the agent travelled down a similar path, which was then blocked by a wall before the agent reached the fork (Figure 3.1b). The wall had two doors, one giving access to each fork. The test pair of objects was shown at the ends of the path, with a red diamond shape at the end of one fork and a blue diamond shape at the end of the other (positions counterbalanced across participants). The video ended with the agent stopping at the fork. For the test pair question participants were then asked “Which door would you like to open?” with the choices “left door” and “right door” positioned under the respective doors. It was predicted that the participants would be motivated to help the agent, and thus the participants’ door choices should indicate which of the two objects they believed the agent to prefer. Participants were then asked “What rule do you think the smiley face was using to choose objects?” with a text box provided for their answer. This awareness question was designed to assess whether they were aware of the rule governing the agent’s choices.

The participants’ answers were coded after the completion of the experiment. Participants’ test pair answers were coded as correct if they chose to open the door leading to the object that matched the colour preference of the agent. Answers to the awareness question were coded as correct if they referenced the colour of the objects rather than other aspects of the choice, such as shape or direction. The test pair question was analysed with a one-tailed binomial test and the answers to the awareness question were presented descriptively. It was expected that participants should be able to determine the colour-based preference rule and therefore should answer the awareness question correctly. If the test pair question effectively prompted participants to answer in line with the agent’s preference this would validate its use in further experiments.

3.2.2 Results & Discussion

For the test pair, the majority (83.0%) of the participants chose the door leading to the object of the preferred colour, and this was significantly more than expected by chance (39/47, one-tailed binomial, $p = 2.77 \times 10^{-6}$). In the awareness question, just three participants failed to report that the agent was making choices based on colour. Five individuals who correctly specified the rule in the awareness question answered incorrectly on the test pair. It is possible that these participants may have misunderstood the test question or may have deliberately chosen not to help the agent; opening the door leading to the non-preferred object. However, overall these results indicate that the test question successfully prompted participants to respond to the agent's preferences and thus validates its use in further experiments.

3.3 Experiment 1

In Experiment 1 the previously validated paradigm was used to investigate adult humans' implicit transitive inference in a 5-term preference hierarchy. In infant studies, shorter hierarchy's were used; 3-term (Gazes et al., 2017; Mascaro & Csibra, 2012; Mou et al., 2014), or 4-term (Mascaro & Csibra, 2014). Using a shorter hierarchy reduces the memory load for the pair-wise choices, but it also means that there is no internal pair that can be tested. All potential transitive pairs in a 3- or 4-term series necessarily involve one of the end terms. Thus, as described above (Section 1.3), the direct history of the end terms can potentially be used to explain any preferences, without employing transitive inference.

As adults' memory capacity is likely to be less limited than infants, I first tested adult humans on a 5-term series. Using a 5-term hierarchy bypasses the end term problem as the internal pair B vs D can be used as the test pair; both elements having been chosen by the agent once and rejected by the agent once so should be equally associated with the agent. Furthermore, as there are more elements to remember, using a 5-term series should minimise the likelihood of the participants determining the hierarchy explicitly before they are asked the test pair question. In order to further ensure that they don't explicitly extract the hierarchy, following the procedure from the dominance-based infant experiments (Gazes et al., 2017; Mascaro & Csibra, 2012, 2014), the participants were presented with each of the 4 pair-wise choices only once before being presented with the test pair question.

3.3.1 Participants

Participants were 292 adults; a further 21 were excluded as during the experiment they did not give full consent for their data to be included in the study. The sample size was determined

through a pilot study as there was no pre-existing literature on the possible effect size for an implicit response to transitive inference in a 5-term hierarchy. It is likely that a larger sample size would be needed than the previous 3-term hierarchy experiments in infants as a 5-term hierarchy involves a greater memory load and may be noisier, meaning that effect sizes may be smaller. The pilot involved 45 participants (with five further participants excluded as they did not fully give consent) and 26/45 answered correctly on the implicit measure. With this effect size, I used G*Power to calculate a minimum target sample size of 245 for a one-tailed binomial. I then rounded this up to 300 to account for exclusions and for potential noise in the pilot study as the estimate of the effect size is likely to be inaccurate (Gelman & Carlin, 2014). Participants were recruited anonymously through Prolific Academic, a website for recruiting study participants, and paid £1 for taking part in the experiment.

3.3.2 Methods & Analysis

The participants were presented with a series of videos through an online survey hosted by Qualtrics. The videos were identical to Experiment 1, except that now each object was a different colour and a different shape: A – purple circle, B – green star, C – orange square, D – blue triangle, E – red diamond. In the training phase, each participant was shown four videos depicting the pairings A>B, B>C, C>D, D>E. The order of presentation of these videos was randomised, and the left/right positions of the two objects in each video were randomised across participants. The original pilot study involved all participants being shown videos representing the same hierarchy. However, for the full experiment the order of the hierarchy was counterbalanced across participants, with 149 of the participants being presented with the hierarchy in the reverse order (E>D>C>B>A). This should control for differences in the participants' responses that are due to differences in the shape or colour of the objects themselves, rather than the preference hierarchy.

The test video was the same as in Experiment 1 - an agent travelling down a forked path that was blocked by a wall (Figure 3.1). The wall had two doors, one giving access to each fork. The test pair of objects was shown at the ends of the path, with object B at the end of one fork and object D at the end of the other (positions counterbalanced across participants). For the test pair question participants were asked, "Which door would you like to open?" with the choices "left door" and "right door" positioned under the respective doors. Participants were then asked an awareness question to assess whether they were aware of the hierarchy governing the agent's choices. The names of the five shapes (purple circle, green star, orange square, blue triangle, red diamond) were presented to the participants in a random order in a 'drag-and-drop' style

question. Participants were asked to “Please rank the objects based on how much you think the smiley face liked them. Most liked objects at the top, least liked objects at the bottom.”

As in the validation, the participants’ answers were coded after the completion of the experiment. Participants’ test pair answers were coded as correct if they chose to open the door leading to object B in the hierarchy. These data were analysed with a one-tailed binomial test. The rankings in response to the awareness question were analysed using the Friedman test to determine if the participants ranked the objects non-randomly, followed by a post-hoc Nemenyi test to determine how the objects’ ranks differed from one another. If the participants are capable of implicit transitive inference then they would be more likely to choose the door leading to B in the test pair question, while their rankings of the objects in the awareness question would be random.

3.3.3 Results & Discussion

In the pilot study, the majority of participants chose the door leading to B (58%). However, this appeared to be caused by an underlying preference for the shape of B (green star) over D (blue triangle). When the hierarchies were counterbalanced for the experiment, so that half of the participants were shown the reversed hierarchy, this pattern disappeared. Participants were equally likely to choose the door leading to B and the door leading to D, and the pattern did not differ from chance (146/292, 50%, one-tailed binomial, $p = 0.523$). The awareness question showed that participants were unable to determine the full hierarchy based on the agent’s pair-wise choices. Out of 292 participants, eleven (3.77%) ranked the full hierarchy correctly and on average participants put 1.38/5 shapes at the correct rank (see Table 3.1). The Friedman test showed that the participants’ rankings were not fully random (Friedman $X^2 = 164$, $p < 2.2 \times 10^{-16}$). However, the post-hoc pair-wise Nemenyi test revealed that this was due to the end term E, which 50.0% of participants placed correctly as the last item of the hierarchy (see Table 3.2). This cannot be explained by the order of presentation of the pairings during training, as the choice videos were presented to participants in a random order.

The above finding is particularly interesting as it resembles the learning pattern often seen in reinforcement-based paradigms with non-human animals. Here, subjects tend to reach a U-shaped accuracy, with a higher level of success on the pairs at each end of the series (A+B- and D+E-) (Vasconcelos, 2008). This phenomenon is referred to as the “serial position effect” (Bryant & Trabasso, 1971; Woocher, Glass, & Holyoak, 1978). Standard serial position effects contrast slightly with the results found here as it was only the least preferred end of the series (E) that was ranked at higher accuracy than the other objects. However, this is not the only example of

such a pattern as both Roberts & Phelps, and Siemann et al. found that accuracy increased the closer a pair was to the non-reinforced end of the hierarchy (Roberts & Phelps, 1994; Siemann, Delius, & Wright, 1996).

Table 3.1 Descriptive statistics for participants' rankings of the five objects in Experiment 1.

	A	B	C	D	E	All
Mean rank	2.87	2.62	2.76	2.71	4.05	3.00
Total Correct	71	71	61	54	146	11
Percentage Correct	24.3	24.3	20.9	18.5	50.0	3.77

Table 3.2 Nemenyi test p-values for participants' rankings of the five objects in Experiment 1, significant p-values are in bold.

	A	B	C	D
B	0.30	-	-	-
C	0.92	0.81	-	-
D	0.72	0.96	0.99	-
E	**3.3×10^{-14}**	**4.7×10^{-14}**	**4.9×10^{-12}**	**3.9×10^{-14}**

3.4 Experiment 2a

In Experiment 1, adults did not show implicit transitive inference on a 5-term preference-based hierarchy after a single presentation of each choice pair. One possible explanation for this result is that the memory requirement for the 5-term hierarchy was too high.

In Experiment 1, I therefore employed a paradigm using 3-term hierarchies that was more similar to the infant experiments (Gazes et al., 2017; Mou et al., 2014). Testing adults on 3-term hierarchies with an intentional action measure could provide further evidence for an implicit sensitivity to transitivity of preferences in humans. Furthermore, as adults can also be questioned explicitly on their understanding of the hierarchy, this allows further investigation of the extent to which logical reasoning underlies such a response. If adults are found to respond

implicitly to the transitivity of the preferences, but are unable to explicitly describe the hierarchy, this finding would indicate that the implicit mechanism is isolated from explicit logical reasoning.

3.4.1 Participants

The previous studies identifying infants' spontaneous transitive inference in a 3-term hierarchy had sample sizes of 32, 24 and 14 infants per condition (Gazes et al., 2017; Mascaro & Csibra, 2014; Mou et al., 2014). As it is possible that the effect size of these studies may be an overestimate (Gelman & Carlin, 2014), a sample size of 50 was chosen in order to exceed the previous experiments. Participants were 49 adults, as one person was excluded because they did not give full consent for their data to be included in the study. Participants were recruited anonymously through Prolific Academic, a website for recruiting study participants online, and paid £1 for taking part in the experiment.

3.4.2 Methods & Analysis

The videos presented to the participants were identical to Experiment 2, except that the hierarchy was reduced from five items to three: A – red diamond, B – blue triangle, C – orange square. Each participant was shown two training videos depicting the pairings A>B and B>C. The order of presentation of these videos was randomised and the left/right positions of the two objects in each video were randomised across participants. The order of the hierarchy was counterbalanced so that 24 of the participants were presented with the hierarchy in the reverse order (C>B>A).

The test video was similar to Experiment 2, except that at the end of one fork was object A and at the end of the other was object C. For the test pair question participants were asked, "Which door would you like to open?" with the choices "left door" and "right door" positioned under the respective doors. Participants were then asked an awareness question to assess whether they were aware of the hierarchy governing the agent's choices. As in Experiment 1, the names of the three shapes (red diamond, blue triangle, orange square) were presented to the participants in a random order in a drag-and-drop question. The analysis of the participants' answers was identical to Experiment 1. If the participants are able to implicitly infer the transitivity of the agent's preferences, they should be more likely to choose the door leading to A than the door leading to C in the test pair question. However, as it is 3-term sequence rather than a 5-term sequence associative explanations would need to be tested in a further experiment.

3.4.3 Results & Discussion

For the test pair, participants were more likely to choose the door leading to A than expected by chance (33/49, 67.3%, one-tailed binomial, p = 0.0107). Participants were also relatively successful at ranking the objects in the 3-item hierarchy, with 20 out of 49 (40.8%) ranking the full hierarchy correctly. On average participants put 1.76/3 shapes at the correct ranking position (see Table 3.3). The Friedman test showed that the ranking of objects was non-random (Friedman $X^2 = 38.3$, $p = 4.76 \times 10^{-9}$). However, again the post-hoc Nemenyi test indicated that this was likely due to correctly ranking the least preferred end term (C), while A and B showed no difference in their rankings between one another (see Table 3.4).

These results indicate that the high level of success on the ranking was not due to an understanding of the full hierarchy, but rather is likely to have been driven by correct positioning of C. This was similar to the result in Experiment 1, with participants in both experiments ranking the hierarchies more accurately than might be expected by chance. In a 5-term hierarchy there are 120 potential combinations (5!), so assuming random ranking, out of 292 participants an average of 2.43 participants would be expected to get the full hierarchy correct. This contrasts with the actual result of 11 participants. Similarly in a 3-term hierarchy, there are 6 potential combinations (3!), meaning that in a sample of 50 an average of 8.33 would be expected to get the hierarchy correct; compared to the much higher obtained number of 20 participants.

These results are somewhat difficult to interpret as it is not clear what is driving this success in ranking the final element in the correct position of the hierarchy. On the one hand, participants could be explicitly able to determine that C or E is the least preferred object in the hierarchy, as comparisons involving non-chosen/non-reinforced stimuli seem to be easiest to recall (see section 4.3). However, on the other hand, it is possible that this placement is driven solely by associative mechanisms, where C is simply least associated with the agent as it has never been chosen and thus is ranked as least preferred. Both A and B have been chosen by the agent once, and thus it may be harder to distinguish between the two.

Table 3.3 Descriptive statistics for participants' rankings of the three objects in Experiment 2a.

	A	B	C	All
Mean rank	1.55	1.73	2.71	2.00
Total Correct	25	22	39	20
Percentage Correct	51.0	44.9	79.6	40.8

Table 3.4 Nemenyi test p-values for participants' rankings of the three objects in Experiment 2a, significant p-values are in bold.

	A	B
B	0.63	-
C	**2.6×10^{-8}**	**3.7×10^{-6}**

3.5 Experiment 2b

In this experiment I investigated further the results of Experiment 2a, assessing whether the adults' response may have been driven by an associative mechanism. In their infant study, Gazes et al. included a control condition in their first experiment. This was designed to test whether the infants were basing their expectations directly on the previous dominance behaviours shown by the puppets (i.e. A is always dominant, C is always subordinate – the end effect), or whether the infants were basing their expectations on transitive inference of the dominance relationship between A and C (Gazes et al., 2017). In this control, after watching dominance interactions for the hierarchy (A>B, B>C), infants watched an end term puppet (either A or C) interact with a novel puppet, D. The puppets were shown either demonstrating a behaviour congruent with their previous behaviour (e.g. A>D) or incongruent with their previous behaviour (e.g. A<D).

If infants' expectations were based only on observed behaviours, they would show a violation of expectation response to the change in behaviour. However, if their expectations were based on transitive inference rather than simply previously observed behaviours, then there would be no prediction about A or C's dominance relative to a novel puppet. Gazes et al. found that infants did not show violation of expectation in the control condition, indicating that the infants may be sensitive to the transitivity of preferences rather than responding based simply on the difference in the puppets behaviour. Here I have adapted this control condition to assess whether the participants in Experiment 2a demonstrated some implicit understanding of transitive inference.

3.5.1 Participants

Participants were 50 adults; one further person was tested but excluded as during the experiment they failed to give full consent for their data to be included in the study. The sample size of 50 was chosen to match Experiment 2a. Participants were recruited anonymously through Prolific Academic, a website for recruiting study participants online, and paid £1 for taking part in the experiment.

3.5.2 Methods & Analysis

The training videos presented to the participants were identical to Experiment 2a. Each participant was shown two videos depicting the pairings A>B and B>C, with A – red diamond, B – blue triangle and C – orange square. The order of presentation of these videos was randomised, and the left/right positions of the two objects in each video were randomised across participants. The order of the hierarchy was counterbalanced so that 24 of the participants were presented with the hierarchy in the reverse order (C>B>A).

The test phase was similar to Experiment 2a except that here each participant saw two test videos. In one video one path led to object A and the other to the novel object D (purple circle), while in the second video the paths led to object C and object D. The participant was asked to indicate which door they would choose to open for the agent, the 'left door' or the 'right door', with the choices positioned under the respective doors. Participants were then asked an awareness question to assess whether they were aware of the hierarchy governing the agent's choices. As in Experiment 2a, the names of the three shapes (red diamond, blue triangle, orange square) were presented to the participants in a random order in a drag-and-drop question.

For each test pair (A vs D and C vs D), one-tailed binomial tests were performed to determine whether participants' choices differed from chance. Furthermore, as each participant was given both questions in a within-subject design, their answers to each test pair were then compared using a McNemar's test for paired data. If the participants were responding based on the statistical association between the agent and the objects they should be more likely to choose the door leading to D when it is paired with C (never chosen) than when it is paired with A (always chosen).

Finally, as in Experiment 2a, the rankings in response to the awareness question were analysed using the Friedman test to determine if the participants ranked the objects non-randomly. Followed by the post-hoc Nemenyi test to determine how the objects' ranks differed from one another.

3.5.3 Results & Discussion

In both of the two test pairings, the participants failed to show a preference for opening one door over the other. In the A vs D pairing, 42.0% of participants chose the door leading to D (21/50, one-tailed binomial, $p = 0.161$), while in the C vs D pairing 54.0% of participants chose the door leading to D (27/50, one-tailed binomial, $p = 0.336$). In addition, the McNemar's test

indicated that the identity of the pairing (A vs D or C vs D) did not affect the participants' likelihood of choosing the door leading to D (n = 50, McNemar's test, $X^2 = 0.593$, p = 0.441).

If the participants were responding only to the agent's choices of the objects, regardless of the hierarchy, then participants would have been more likely to pick D when it was paired with C than when it was paired with A. As this was not the finding here, these results suggest that the participants' responses in Experiment 2a were unlikely to be based only on the statistical association between the agent and the object. Thus, this indicates that in Experiment 2a participants may have been sensitive to the transitivity of the relationship between A and C.

Participants' success on the awareness question appeared to be lower than in Experiment 2a, with 11 out of 49 (22%) ranking the full hierarchy correctly, and on average participants put 1.3/3 shapes at the correct position (see Table 3.5). This may be because the participants were asked two test questions before the ranking rather than one, prolonging the delay between presentation of the agents' choices and the ranking question. In addition, the implicit questions in Experiment 2b also introduced a novel stimulus, D, which may have interfered with participants' recollection. The Friedman test showed that the ranking of objects was non-random (Friedman $X^2 = 8.92$, p = 0.0116). The Nemenyi test indicated that this was due to correctly ranking the non-chosen end term C. However, in contrast to Experiment 2a, while C was different to B there was no difference between C and A, although the p-value was 0.056 which may suggest a trend (see Table 3.6).

Gazes et al. also added a further experiment to their study to clarify whether the infants' response were based on transitive inference. The authors suggested that their results could be due to the incongruent C>A video involving two reversals of the dominance shown in the training videos (both A and C), whereas the incongruent control only involved one reversal (either A or C). Their original control condition was similar to that used here, comparing looking times between congruent videos (A>D or D>C) and incongruent videos (D>A or C>D). However, the infants were only presented with the two videos featuring A or the two videos featuring C. This contrasts with the experiment presented here as the participants were questioned on both A vs D and C vs D in a within-subject design. In this experiment, I compared participants' answers between the two conditions, accounting for the reversal of both previously viewed preferences. Thus a follow up experiment comparing a double reversal, as in Gazes et al.'s study, is not necessary here (Gazes et al., 2017).

Table 3.5 Descriptive statistics for participants' rankings of the three objects in Experiment 2b.

	A	B	C	All
Mean rank	1.88	1.78	2.34	2.00
Total Correct	21	19	26	11
Percentage Correct	42.0	38.0	52.0	22.0

Table 3.6 Nemenyi test p-values for participants' rankings of the three objects in Experiment 2b, significant p-values are in bold.

	A	B
B	0.871	-
C	0.056	**0.014**

3.6 General Discussion

The experiments in this chapter were designed to add to the investigation of an implicit sensitivity to the desires of others. Specifically, the experiments extended previous findings in infants that indicate sensitivity to the transitivity of others' preferences. I designed a novel paradigm and validated its use with a simple preference rule, identifying that the test question was able to prompt participants to respond to the agent's preferences. I then employed this paradigm in three experiments assessing adults' implicit sensitivity to the transitivity of a preference hierarchy.

In Experiment 1, I found no evidence for transitive inference of preferences when adults were presented with a 5-term hierarchy. A likely explanation for this is that the participants were unable to remember the four pair-wise choices made by the agent, either implicitly or explicitly. Although their ranking of the objects was non-random, this was not due to construction of the full hierarchy but was instead driven by an increased probability of a low ranking for the final element in the hierarchy.

In Experiment 2a I employed a shorter 3-term hierarchy instead of a 5-term hierarchy, matching the previous experiments done with infants. Here, on the test pair question participants were more likely than expected by chance to choose the door leading to the preferred object. However, the results of the awareness question complicate the interpretation of this result. As

in Experiment 1, participants' rankings were non-random, with the least preferred object, C, being ranked significantly lower than the other objects. As this was one of the objects used in the test pair it is possible that the success on the test pair may have been driven by the explicit understanding that C is the least preferred object in the hierarchy.

Furthermore, as the hierarchy only contained 3 terms, the test pair necessarily contained the end terms: A which was always chosen, and C which was never chosen. To address the possibility of an end effect, in Experiment 2b I tested whether the agent's choice of the objects was sufficient to produce the participants' response to the test pair question in the absence of a hierarchy. Participants did not choose the door leading to the preferred object more often than expected by chance. Thus, an associative explanation does not appear to explain the results found in Experiment 2a, supporting the suggestion that there may be some sort of implicit sensitivity to transitive inference shared between human infants and adults.

This result contrasts with previous findings that have suggested adults' implicit responses to reinforcement-based transitive hierarchies may be explained by associative mechanisms (Frank et al., 2005). These previous experiments involved extensive training on the pairs, and participants were provided with feedback on whether their answers were correct or incorrect. Thus, the stimuli were differentially rewarded and participants had ample opportunity to develop differing associations between the stimuli and a reward.

This contrasts with the paradigm I used here in two ways. Firstly, the choices were made by the agent rather than the participant and the participant was therefore unrewarded. Secondly, the participants were only shown the agent making each choice once, making it less likely for an association to develop. These factors may have contributed to the lack of an associative explanation, or end effect, found in Experiment 2b. Similarly, in the infant transitive inference studies, the authors also included control conditions that failed to find plausible associative explanations for their results (Gazes et al., 2017). Simple associations may therefore be less likely to explain transitive inference in a social paradigm (dominance- or preference-based) than in a traditional reinforcement-based paradigm.

However, this must be considered with the caveat that these infant studies have low sample sizes and noisy data; and that, although there are some consistencies between the studies, none of them directly replicate the findings of each other(Gazes et al., 2017; Mascaro & Csibra, 2012, 2014; Mou et al., 2014). As such, this field may suffer from similar issues to those currently seen in false belief studies employing looking measures (Kulke & Rakoczy, 2018; Poulin-Dubois et al., 2018) (see Chapter 2), and would benefit from large-scale direct replications.

Furthermore, there are also arguments over the interpretation of violation of expectation measures used in the infant preference study (Mou et al., 2014). Where some argue that such spontaneous responses reflect an understanding of theory of mind (Baillargeon et al., 2010; Carruthers, 2013) or a distinct implicit system (Apperly & Butterfill, 2009; Csibra & Gergely, 1998; Nichols & Stich, 2003; Ruffman et al., 2012), others argue that they are simply a consequence of low level novelty or other artefacts (Heyes, 2014a). It is therefore difficult to determine what level of cognition might underlie both infants' and adults' implicit responses to preference-based hierarchies.

Finally, across experiments the results from the awareness questions were less clear cut than might have been desired. As participants were able to rank the final term in the correct position, it is possible that their correct answers to the test pair question may have involved accessing explicit knowledge about object C being the agent's least preferred object. The Nemenyi test in Experiment 2a found a difference between participants' ranking of objects A and C ($p = 2.6 \times 10^{-8}$), but no such difference was found in Experiment 2b ($p = 0.056$). Thus, it is possible that the difference in the intentional action responses (choice of door) in the experiments may have resulted from explicit reasoning about the hierarchy.

In summary, I found that adult humans did not appear to respond implicitly to an agent's 5-term preference-based hierarchy, but did respond implicitly to a 3-term hierarchy. Furthermore, I found no evidence that these responses were driven by associative processes. However, it is unclear to what extent these results might be driven by a partial explicit representation of the hierarchy. Overall, this chapter highlights the need for a broader investigation of transitive inference beyond the traditional reinforcement-based paradigm, particularly focussing on the social domain.

4. Chapter 4 – Eurasian jays' attribution of specific satiety in co-operative contexts

4.1 Introduction

During breeding season male Eurasian jays are highly motivated to feed larvae to their female partners in direct beak-to-beak exchanges. Studies investigating this behaviour have found that when their partner was fed to satiety on a specific food, male Eurasian jays were able to alter their pattern of food sharing in line with their partner's reduced desire for that food (Ostojić et al., 2014, 2013; Ostojić, Legg, et al., 2016). However, it is currently unclear whether this response results from a generalised sensitivity to their partner's desires, or a specialised ability related to food sharing (see also Chapter 5).

Adaptive specialisations are traits developed by a species in response to a specific ecological challenge. This contrasts with generalised traits or abilities which can be applied in a range of scenarios. An often proposed cognitive adaptive specialisation is the use of spatial memory by food-storing birds, as they have an enlarged hippocampus relative to body size (Krebs, 1990; Shettleworth, 1990)[11]. Given the likely high level of natural selection on courtship behaviours like food sharing, it is possible that male Eurasian jays may also possess a cognitive adaptive specialisation involved in food sharing that allows them to respond to their partner's desires. One way to distinguish between an adaptive specialisation and a generalised theory of mind might be to design novel paradigms that could assess jays' sensitivity to their partner's specific satiety outside of direct beak-to-beak food sharing.

Furthermore, novel paradigms might allow the expansion of this investigation in new directions beyond the limitations of the existing paradigm. The current food sharing paradigm is limited to the time of year when food sharing is most frequent (breeding season), is limited to testing males (as females rarely share food), and can only assess corvid species that carry out food sharing behaviours. There are large seasonal changes in brain structures associated both with bird song (Nottebohm, 1981) and food caching (Smulders, Sasson, & DeVoogd, 1995), and there may also be changes in the brain that arise from seasonal variation in social complexity (Emery, Seed, et al., 2007). Novel paradigms that do not require food sharing could be used to test whether social cognitive abilities vary throughout the year, whether jays are able to respond to

[11] This view is not without dissenters. In their review of the evidence for an adaptive specialisation of memory in food-storing birds, Macphail and Bolhuis argue that hippocampal enlargement alone is insufficient evidence for an adaptive specialisation, and disagree with the conclusion that storers have superior spatial memory to non-storers (M. Macphail & J. Bolhuis, 2001).

desire changes in non-partners, and whether there are differences between males' and females' abilities. Thus, in this chapter I have explored alternative paradigms that might allow investigation of Eurasian jays' sensitivity to their partner's specific satiety without using beak-to-beak food sharing.

4.2 Experiment 1

In Chapters 2 and 3 I discussed some of the looking measures used to investigate implicit theory of mind in human adults and infants, namely anticipatory looking and violation of expectation. One piece of evidence that supports a potential shared minimal theory of mind between humans and non-human animals is the demonstration of anticipatory looking responses in apes (Krupenye et al., 2016; Krupenye, Kano, Hirata, Call, & Tomasello, 2017). However, eye tracking measures have not been traditionally used with birds as the position of their eyes allows them to switch between frontal (binocular) and lateral (monocular) vision, making it harder to determine what they are focusing on (Stamp Dawkins, 2002).

Despite this limitation, studies of birds looking preferences have been conducted by using peepholes to restrict their looking behaviour. In order to study jackdaws' and ravens' attention to conspecifics, Scheid et al. used a paradigm based on a previous study involving marmosets, in which visual access was restricted to two small holes in an opaque partition (Range & Huber, 2007; Scheid, Range, & Bugnyar, 2007). Both of the holes looked onto the same compartment, providing slightly different viewing angles. This paradigm was subsequently modified by Bird and Emery, such that each of the two peepholes gave visual access to a different compartment, allowing ravens' looking preference to be determined between two choices (Bird & Emery, 2008). This basic two-peephole premise has been employed by a number of subsequent studies employing looking measures with birds (Grodzinski, Watanabe, & Clayton, 2012; Shaw & Clayton, 2013; Watanabe, Grodzinski, & Clayton, 2014), including violation of expectation paradigms investigating birds' understanding of support relationships (Bird & Emery, 2010; Davidson et al., 2017). Here I have attempted to adapt this paradigm to see if it could be used to assess Eurasian jays' sensitivity to others' specific satiety.

It has been shown in previous experiments that corvids' have a propensity for watching conspecifics, particularly their partners (Bird & Emery, 2008; Scheid et al., 2007). There is also some evidence that scrub jays will check back on locations of interest (caching trays) to update their knowledge (Grodzinski et al., 2012). It is therefore predicted that a jay would be motivated to look into the compartment they anticipate another bird entering, especially if it is their partner. Such a paradigm could then be used to assess an understanding of their partner's

specific satiety by having each peephole looking into compartments containing different food types; the food that the partner was sated on and a novel food type. If the observer has an understanding of their partner's satiety they should anticipate their partner approaching the non-sated food and therefore look preferentially into that compartment.

Peepholes have been used previously with Eurasian jays (Davidson et al., 2017; Shaw & Clayton, 2013) and it was therefore expected that this paradigm should be appropriate to use with these birds. However, it was possible that jays may not be motivated to look into the compartment they expect their partner to enter. As such, before testing the birds on the specific satiety paradigm I conducted a validation experiment. This involved altering their partner's physical access to the compartments (Figure 3.1). If jays are able to predict their partner's location based on physical access they should demonstrate preferential looking towards an open compartment over a closed compartment. This should be an easier task than the attribution of specific satiety to their partner. If the birds show preferential looking in this simpler version of the task, this would then validate the experimental set-up and allow it to be used for investigating an understanding of specific satiety. To facilitate the birds' ability to determine the location their partner had access to, the birds were given personal experience of the apparatus, as well as being given experience of watching their partner enter the compartments.

4.2.1 Subjects

Birds from five male-female pairs from two colonies (Colony 1 = 3 pairs, Colony 2 = 2 pairs) were involved in this experiment. This consisted of all living pairs from Colony 1, and two pairs from Colony 2 chosen by the experimenter Ljerka Ostojić. From these pairs seven birds were used as observers (3 male, 4 female). The remaining three birds were excluded as observers (Caracas, Lima and Hunter); one because they failed the second familiarisation stage (Hunter), and two because their partners did not pass the first familiarisation stage (Caracas and Lima). The experiment was conducted from February to June 2017 and all test trials occurred between paired birds during the breeding season. Birds were housed in outdoor aviaries (20 × 6 × 3m) and tested in indoor compartments (2 × 1 × 2m). Birds were fed a maintenance diet of soaked cat biscuits, vegetables, seeds and fruit and had ad libitum access to water. To ensure that the birds were mildly hungry and thus motivated to act for food, maintenance diet was removed from the aviary approximately 1 hour before the start of testing. The birds were tested once per day between the hours of 9am and 12pm, and the timing of testing for each pair was kept as similar as possible between days to keep their motivation consistent.

4.2.2 Methods

4.2.2.1 Apparatus

The experimental set-up consisted of three separate indoor experimental compartments (Figure 3.1). The observing bird was in the far right compartment, while the actor (their partner) would start from the far left compartment. These locations could not be counterbalanced as the set-up involved a fixed box put in place to modify the structure of the compartments. The central compartment was divided in two, and access to the two sections could be controlled using a sliding door. Each section of the divided compartment had a small metal bowl containing waxworms (35g, approx. 200). Food was always present in both compartments, even though only one was accessible to the actor. This should prevent the observer's looking behaviour from being influenced simply by the presence or absence of food. The actor could be let through into the central compartment by removing a transparent barrier between the two compartments (dashed line). The experimenter always left the room after the transparent barrier was removed so that their presence would not influence the looking behaviour of the observer. The wall between the central compartment and right compartment had two peepholes with individual perches, as well as a large window above the peepholes that could be opened or closed using a blind.

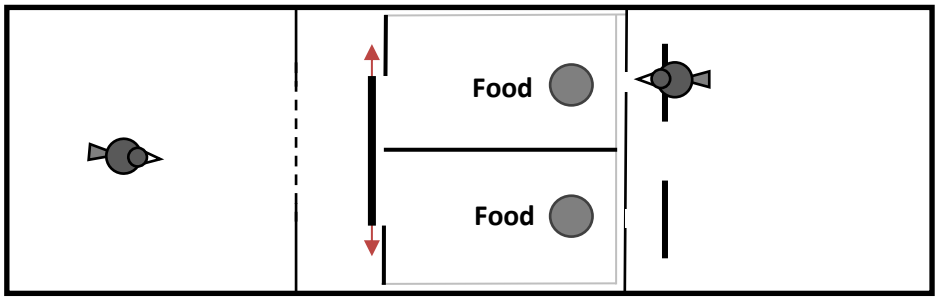

Figure 3.1 Plan view illustrating the experimental set-up used to determine whether Eurasian jays will preferentially look towards the compartment that their partner has physical access to. The observer was in the right compartment, while the actor started in the far left compartment. The dashed line represents a transparent barrier that could be removed to allow the actor access into the central compartment. The central compartment contained a wooden-framed mesh box divided into two sections. Access to the two sections could be controlled by sliding an opaque door (thick black line, red arrows indicate direction of movement). Each section contained an identical bowl of food. The observer's compartment contained two perches with peepholes looking into each section of the central compartment.

4.2.2.2 Familiarisation

Birds were given two familiarisation stages, in which they experienced the experimental set-up as both actors and observers. The actor familiarisation stage involved two trials, one trial in which the sliding door was to the far left so that the right section was open and one trial in which the sliding door was to the far right so that the left compartment was open. The order of the two trials was counterbalanced across individuals. A bird passed this familiarisation stage if they were willing to approach the food in both sides of divided compartment within 3 minutes of the transparent barrier being removed. If the actor failed to enter within 3 minutes on the first trial, the trial was repeated once more on a subsequent day, for a maximum of two trials on each side. If the bird failed to approach the food within 3 minutes the bird was excluded as an actor and thus their partner was not used as an observer.

In the observer familiarisation stage each bird was given the opportunity to watch their partner while they ate the food from the central compartment, alternating between the left and right side being open. The order of trials was counterbalanced across observers. In the first two of these familiarisation trials the blind over the window was left open so that the observing bird had a full view of their partner. In subsequent trials, the blind over the observer's window was closed so they could only see into the divided compartment through the peepholes. This familiarisation was required to give the observer experience of the actor entering the open section, so they should be able to predict which section the eater will enter in the test trials.

The looking behaviour of the observer was recorded using a camera for 10 minutes following the removal of the transparent barrier. The observer was considered to have passed the familiarisation stage if they looked through the peepholes in two consecutive blind-closed trials. The actor was also required to enter the divided compartment, to ensure that the observer had experienced looking through the peephole when the actor had gone through. Blind-closed trials were conducted up to three times on each side (a maximum of six days in total). If the observer had not looked through the peephole on two consecutive trials by the final day they were excluded from acting as observers in the test trials. However, they could still be used as an actor in the test trials if they had passed the actor familiarisation stage.

4.2.2.3 Testing

The test phase of the study involved two conditions in a within-subject design: a social condition in which their partner was present and a non-social control in which their partner was absent. In the social condition two identical transparent barriers were placed over the hole between the left compartment and the central divided compartment. Once the observer and actor were in

place they were left for 5 minutes before the blind was closed to ensure that the observer knew that their partner was present. The sliding door was then moved into either the far-left or far-right position, and one of the transparent barriers was removed. The second barrier was still in place so the actor was not actually able to enter the divided compartment. However, from the perspective of the observer this action should look identical to the familiarisation trials, and thus they should anticipate that their partner would enter the open section. The looking behaviour of the observer was recorded using a camera for 10 minutes following the removal of the transparent barrier.

The non-social condition was identical except that the observer's partner remained in the aviary while the test trial was conducted. If the males show a preference for looking into the open compartment when their partner is present but not when their partner is absent, this would indicate that the looking preference is driven by looking into the compartment that their partner has access to, rather than due to any physical differences in the scene (i.e. the open door).

Each bird experienced four test trials: a social trial with the left section open, a social trial with the right section open, a non-social trial with the left section open, and a non-social trial with the right section open. If a bird failed to look through either of the peepholes in a trial, a repeat of that trial was carried out after all other test trials had been completed (Bird & Emery, 2008; Scheid et al., 2007). If the bird did not look through either peephole again in the repeated trial, their looking times were recorded as zeroes for this trial. The test trials were interleaved with further familiarisation trials in which the eater was let through to the food, to preserve the observer's interest in looking through the peepholes. All trials alternated between left and right sections being open and the order of trials was counterbalanced between birds.

The looking times were measured by coding the video recordings after the trials had concluded. The videos were exported into VLC player and looking behaviour was recorded through experimenter observation. The birds were considered to be looking through a peephole when they stood on the appropriate perch and engaged in 'peering' behaviour (Bird & Emery, 2008). The looking times for each peephole were summed across the left and right trials, so that the total time spent looking into the open section and the total time spent looking into the closed section were determined. This should avoid any differences in looking that are driven only by a preference for one side over the other.

4.2.3 Analysis

A Wilcoxon signed-rank test was used to compare the looking times between the social and non-social conditions. This should indicate whether the time spent looking through the peepholes

was influenced by the presence of their partner. A further Wilcoxon signed-rank test was then used to determine whether the birds showed a different looking pattern between the social and non-social conditions. If the observer anticipated their partner entering the open section, they should spend a greater proportion of total looking time looking into the open section when their partner was present (social condition) than when their partner was absent (non-social condition)[12].

4.2.4 Results & Discussion

The jays spent longer looking through the peepholes in the social condition than the non-social condition (two-tailed Wilcoxon, n = 7, V = 28, p = 0.0156) (Figure 3.2), suggesting, in line with previous studies (Bird & Emery, 2008; Scheid et al., 2007), that the observers showed greater interest in looking through the peepholes when their partner was present then when their partner was absent. However, in the crucial comparison there was no difference in the proportion of time spent looking into the open section between the social and non-social conditions (two-tailed Wilcoxon, n = 7, V = 10, p = 0.578) (Figures 3.2 and 3.3).

Thus, I found no evidence that the birds anticipated which section their partner was likely to enter. However, this may not have been due to a lack of understanding of physical access, but rather due to the birds' lack of motivation to look through the peepholes. Six out of seven birds needed a repeat of at least one of the trials as they failed to look through either peephole within the 10 minute recording period; three of these birds repeated two trials and one bird repeated three trials. During these repeats, four out of seven birds showed no looking in at least one condition, with one bird failing to look in repeated trials across two of the four conditions. This indicates a general lack of interest in the peepholes and low motivation to look through them.

[12] The proportion of time spent looking into the open section was derived using the formula $\frac{O}{O+C}$ where O is the total time spent looking into the open section, across the left and right trials, and C is the total time spent looking into the closed section. The proportion of total looking time spent looking into the closed section is the inverse of this value and therefore does not alter the output of the analysis.

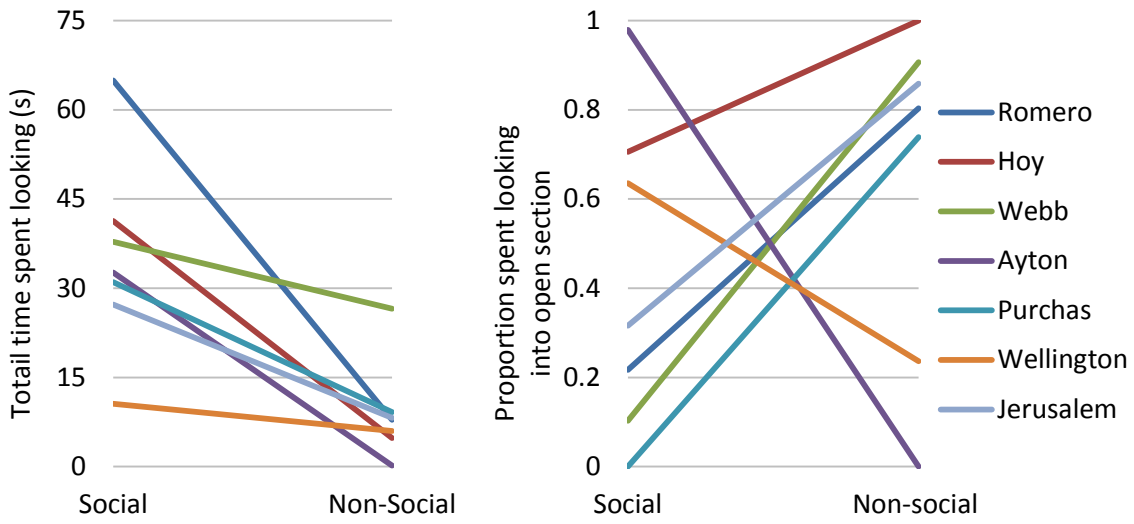

Figure 3.2 The left panel shows the difference in the total time spent looking through the peepholes between the social and the non-social conditions. The right panel shows the difference in the proportion of time spent looking into the open section between the social and non-social conditions. Paired data from each bird are connected by coloured lines to show the direction of the difference between conditions. If birds' looking preference was related to their partner's ability to access the section, the proportion of time spent looking into the open section should be higher in the social condition than the non-social condition.

Previous experiments using looking time measures in corvids have also had some difficulties with the birds' low motivation to look through the peepholes. In their study Scheid et al. repeated their initial experimental sessions for those birds that failed to approach the holes (5 individuals), and Bird and Emery repeated trials in which their birds failed to approach the holes within 15 minutes (Bird & Emery, 2008; Scheid et al., 2007). However, in these studies the birds became familiarised with looking through the peepholes over time or looked through the holes on subsequent trials. In my experiment, the recording time of 10 minutes was slightly shorter than that use in other experiments involving a choice between two different holes (Bird & Emery, 2008; Grodzinski et al., 2012; Shaw & Clayton, 2013; Watanabe et al., 2014). However, this shorter recording time was a deliberate decision as during breeding season Eurasian jays show distress when separated from their partners for prolonged periods of time. The birds were tested during breeding season to ensure that they would have the greatest motivation to look for their partner, however, it is possible that being separated from their partner may have altered their behaviour in such a way that it reduced their motivation to look through the peepholes.

In addition, it is also possible that any preference for looking into the open compartment may have been masked by the extensive side biases displayed by the birds. Although the jays spent a greater time looking through the peephole in the social condition than the non-social condition, almost all of this extra looking time appeared to be through the left peephole (Figure 3.3). This side bias is not a unique feature of this paradigm but rather a response that has been found in a number of scenarios (see Chapter 6). This could be a consequence of various factors such as previous experiments in which certain locations were differentially rewarded, the layout of the compartment set-up or the inherent laterality of certain behaviours (Martinho, Burns, von Bayern, & Kacelnik, 2014).

Overall, the paradigm I trialled here was found to be ineffective for investigating Eurasian jays' senstivity to their partners' specific satiety because the birds were not sufficiently motivated to look through the peepholes. Furthermore, when they did look through the peepholes this action appeared to be unrelated to their partners' access to the two sections. Therefore, in the next section I describe a paradigm designed such that the jays should have a greater motiviation to produce the necessary behaviour.

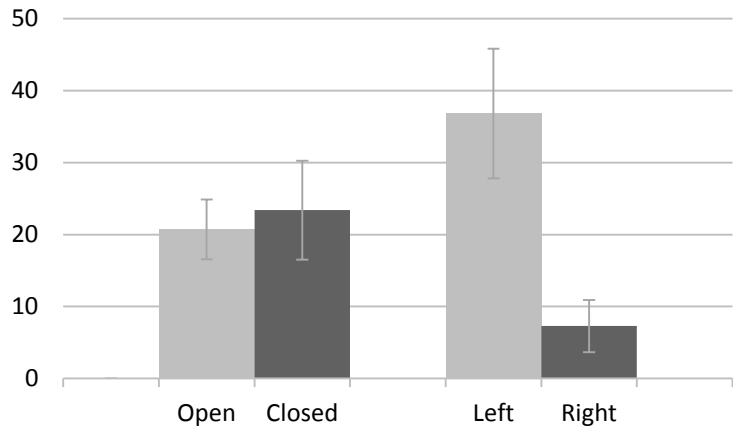

Figure 3.3 An illustration of the effect of the birds' side biases on their looking patterns. The left hand columns represent the mean time spent looking through the peepholes into the open section (light grey) and the closed section (dark grey), for all birds across trials. The right hand columns represent the mean time spent looking through the left peephole (light grey) and the right peephole (dark grey), for all birds across trials. Error bars show standard error.

4.3 Experiment 2

A number of previous studies have demonstrated that male jays respond to the specific satiety of their partner when engaged in direct beak-to-beak food sharing (Ostojić et al., 2014, 2013;

Ostojić, Legg, et al., 2016). Therefore, in Experiment 2 I attempted to address the motivational problems found in Experiment 1 by developing a paradigm that involved jays actively provisioning their partners with food. Including a food provisioning component might help to maintain the birds' motivation to express the behaviour of interest, as well as encouraging them to track their partners' behaviour. The paradigm tested here involved a pro-social apparatus, in which a bird could pull a string that turned an arm to deliver food to their partner. The male should be motivated to provide the food to the female during breeding season as the paradigm mirrors their natural food sharing behaviour. Furthermore, there has been some past success with corvids on pro-social paradigms, in which they help others at no cost to themselves. There is evidence that both jackdaws (Schwab, Swoboda, Kotrschal, & Bugnyar, 2012), and azure-winged magpies (Horn, Scheer, Bugnyar, & Massen, 2016) are willing to provide food to conspecifics outside of a courtship food sharing context.

4.3.1 Subjects

Four male-female pairs from two colonies (Colony 1 = 2 pairs, Colony 2 = 2 pairs) were tested during this experiment, with the males as actors and the females as receivers. These four males were chosen as they were the only paired individuals that would pull the strings on the boxes; one female (Washington) was also able to open the boxes; however she was excluded from testing as she had no partner. Familiarisation and training with the apparatus was conducted between November and December 2017 for Colony 1 and between February and March 2018 for Colony 2. The test trials were conducted during the breeding season for both colonies in April and May 2018. The birds were housed in outdoor aviaries (20 × 6 × 3m) and tested in indoor compartments (2 × 1 × 2m). They were fed a maintenance diet of soaked cat biscuits, vegetables, seeds and fruit and had ad libitum access to water. To ensure that the birds were mildly hungry and thus motivated to act for food, maintenance diet was removed from the aviary approximately 1 hour before the start of testing. The birds were tested once per day between the hours of 9am and 12pm, and the timing of testing for each pair was kept as similar as possible between days to keep their motivation consistent.

4.3.2 Methods & Analysis

4.3.2.1 Apparatus

For this experiment, I designed an apparatus that allowed the male jay to provide their partner with food, without directly passing it from beak to beak. This involved a box structure with a rotating arm that could be used to deliver food to a partner on the other side of a mesh barrier

(Figure 4.1). As the male pulled the string one bowl moved towards him into his compartment, while the other bowl swung away into the adjacent compartment. When the arm reached a perpendicular position it hit a stop with a small magnet that held the arm in place and prevented it from swinging. Using two of these boxes it should be possible to test the males' sensitivity to their partner's specific satiety. If a waxworm is placed in one of the boxes and a mealworm in the other, the box the male chooses to open should indicate which type of worm he prefers to provide to the female. This could be paired with pre-feeding of the female to investigate whether the male's food provision choices were affected by the female's satiety.

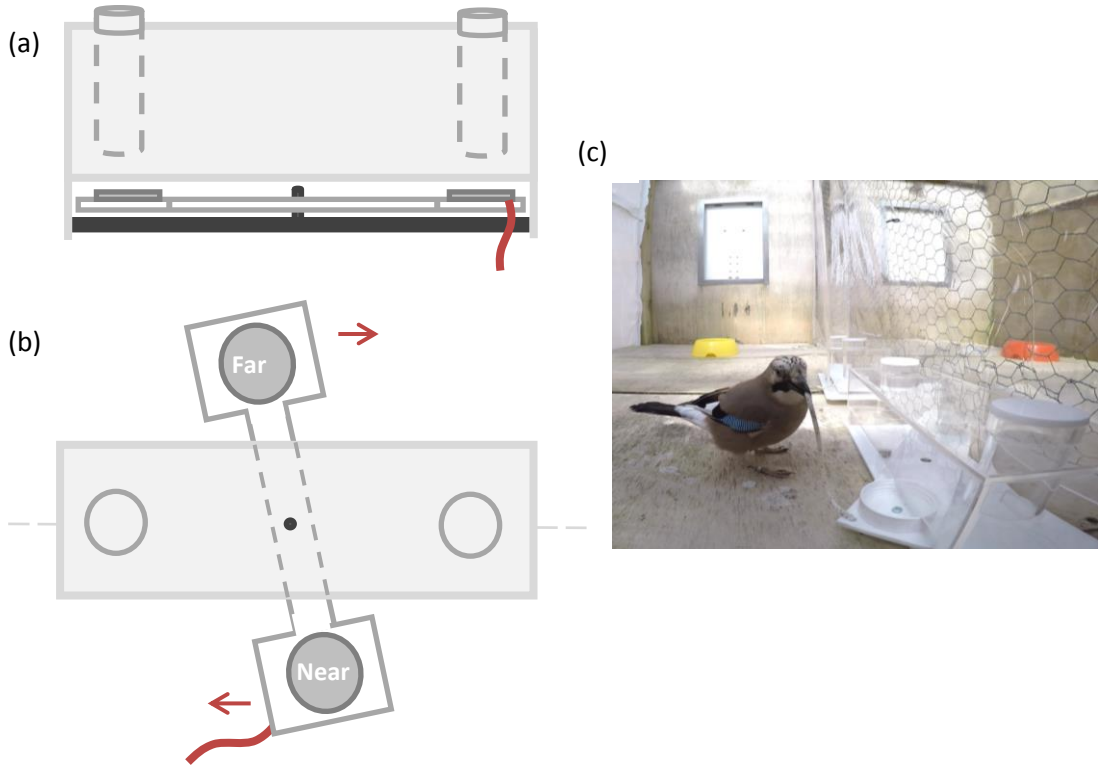

Figure 4.1 Experimental apparatus used in Experiment 2; (a) side view, (b) plan view. The boxes were made of clear Perspex with an open slit at the bottom on each side with enough clearance for the spinning arm to pass through. The boxes also had clear tubes to allow the food bowls to be refilled without pulling the arm out. The tubes could be covered by opaque plastic lids during testing. The arms were made of white plastic with a small white bowl attached at each end. The arms had a small piece of white string (shown in red) attached to the near end, which could be used to pull the arm out from under the box. The red arrow indicates the direction of movement of the arm when the string was pulled. A small magnetic stop inside the box prevented the arm from being pulled past the perpendicular position, and held the arm in position so it wouldn't swing while birds were eating from the bowls. (c) shows a still from a test trial video, in which the spinning box apparatus is being opened by one of the male subjects (Romero).

4.3.2.2 Familiarisation & Training

The males were first allowed to interact with the apparatus in the main aviary until they consistently approached the box and pulled the string to gain food rewards. Subsequently, the four males that spontaneously interacted with the box received a series of training stages to ensure that they were paying attention to the content of the bowls in the box, and not just pulling the strings due to their association with a food reward. Together these training stages should encourage the males to pay attention to both bowls and give them experience of the far bowl being accessible from the female's adjacent compartment. The males were given up to ten training trials a day. If the birds showed any sign of distress they were released early from the testing compartment. To pass each training stage and move on to the next stage the males had to perform the correct behaviour on three consecutive trials. If the male performed incorrectly on three consecutive trials they were moved back to the previous training stage.

Stage 1a: In the first stage of the training two boxes were placed in the male's compartment, one with food placed in both bowls (baited), and one that was empty (unbaited). In order to access the food and pass the task, the male had to pull the string of the baited box. As soon as the male pulled the string of one box, the other box was removed, so that if the male approached the unbaited box first he did not get a food reward. The position of the baited box was pseudo-randomised such that the baited box was never in the same location for more than two consecutive trials. The boxes were placed centrally in the male's compartment so that the male had access to both the bowls when the arm was turned (Figure 4.2a). Each trial was 2 minutes long, and if a bird failed to act on the apparatus within 2 minutes the boxes were removed and a new trial started. To pass this stage the male had to pull the string of the correct box and take the food in both bowls on three consecutive trials.

Stage 1b: In the next stage, the same set-up was then repeated but with the boxes placed against the mesh of the male's compartment. The mesh had slits cut into it, which allowed the arm to swing through so that when the string was pulled the far bowl would end up on the other side of the mesh (Figure 4.2b). The slits in the mesh were not large enough to allow the birds to pass through. A larger opening in the mesh beside the boxes gave the male access to the second compartment. The position of the baited box was pseudo-randomised between trials. The males were required to interact with their chosen box within 2 minutes, but were then given an extra minute to go through to the neighbouring compartment, giving a total trial time of 3 minutes. To

pass this stage the male had to pull the string of the correct box and go through the opening to the other compartment to access the food on three consecutive trials.

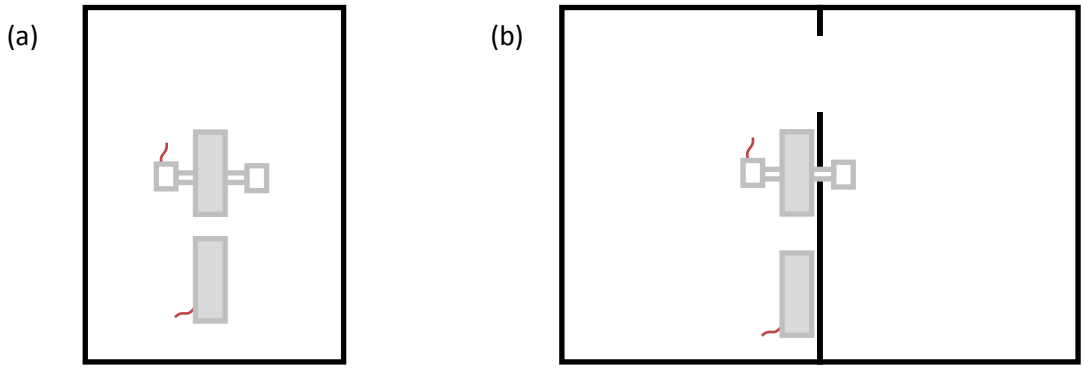

Figure 4.2 Illustration of the compartment set-up used in the training stages of Experiment 2. The boxes are shown in grey, with the string in red, and with the top box in each picture showing the position of the arm when the box is opened. In stages 1a and 2a the boxes were placed in the middle of the male's compartment (a). In stages 1b and 2b the boxes were pushed up against the mesh between two adjacent compartments (b). A slit was cut in the mesh that was large enough to allow the arm to swing through but not large enough for the birds to move through. A further hole was cut in the mesh next to the boxes (gap in the black line) to allow the male bird to travel into the adjacent compartment and get the food from the far bowl.

Stage 2a:[13] This training stage was similar to 1a except that food was placed only in the far bowl of one box (baited) while the other box was empty (unbaited). To pass the task, the male had to pull the string of the baited box. As soon as the male pulled the string of one box, the other box was removed. The position of the baited box was pseudo-randomised between trials. The boxes were placed centrally in the male's compartment, and each trial was 2 minutes long. To pass this stage the male had to pull the string of the correct box and take the food in the bowl on three consecutive trials.

[13] I initially planned an alternative second familiarisation phase to train males to pay attention to both bowls of the box (both near and far). Food was placed only in the near bowl of one box, while food was placed in both bowls of the other box; the males were expected to choose the box containing more food. However, in preliminary investigations I found that when food was present in the near bowls, the jays did not appear to attend to what was in the far bowls. The birds did not show a preference for the box containing more food, tending to approach whichever box they were closest to at the start of the trial. This may have been caused by an inability to inhibit a prepotent response for pulling the string near to the food.

Stage 2b: Stage 2a was then repeated with the boxes up against the mesh, similarly to Stage 1b. An opening in the mesh beside the boxes gave the male access to the second compartment. The trial time was 3 minutes, giving an extra minute for the birds to move through into the other compartment and retrieve the food. To pass this stage the male had to pull the string of the correct box and go through the opening to the other compartment to access the food on three consecutive trials.

Stage 3: The final stage of training required the males to pass a task involving two different foods. Initially the plan was to test the males on an "own specific satiety" paradigm, where they had to choose the box that aligned with their own satiety to a pre-fed food. However, I found that the males would not pull the strings of the boxes when they were already sated, likely due to a decreased interest in the food reward.

Instead the males were given a preference task, in which different foods were placed in the far bowl of each of the two boxes. If the males were able to attend to the boxes and distinguish between the types of food, their choices should match with their food preference. The males' preference within each food pair was tested by giving them a series of choices between the two foods presented in normal bowls. The bowl preferences and box preferences were recorded on different days and their order counterbalanced. To pass the preference test the birds had to show alignment of their preferences for a minimum of 4 food combinations, and were tested on a maximum of 5 combinations. This preference test should validate the procedure as it indicates that the males are attending to the type of food present in the far bowls, and are able to successfully make choices between them.

Stage 4: Once the males had passed the training trials and the preference test, their female partners had to be familiarised with the boxes so that they would eat from the bowl. Before conducting the test trials all the female partners were familiarised with the boxes such that they were comfortable with approaching the box within 2 minutes and taking the food from the bowls.

4.3.2.3 Testing

During testing the opening between the two compartments was closed using a transparent plastic barrier, and the male and female were in adjacent compartments. In all test conditions the female was pre-fed for 15 minutes while the male was present. In the baseline the female

was given a handful of maintenance diet, while in the specific satiety conditions the female was given either 50 waxworms or 50 mealworms. After pre-feeding the male received 10 test trials lasting 3 minutes each. For each trial, two boxes were placed in the male's compartment against the mesh of the female's compartment, one box had a mealworm in its far bowl and the other box had a waxworm in its far bowl. No food was placed in the near bowls for the males as this may have distracted them from attending to the far bowl (see Footnote 14).

In each trial, as soon as the male interacted with the string of one box the other box was removed so that he was only able to provide a maximum of one worm to the female per trial. Each of the three conditions: baseline, pre-fed waxworms, and pre-fed mealworms, were conducted twice on consecutive days to give a total number of 20 trials per condition, making it comparable to the 20 choices presented to the male in food sharing studies (Ostojić et al., 2014, 2013; Ostojić, Legg, et al., 2016). The baseline was always performed first, while the order of the two specific satiety conditions was counterbalanced between subjects. Given that only four birds were familiarised and trained on the apparatus, all results are reported descriptively.

4.3.3 Results & Discussion

The four male birds took between 4 and 10 days to pass the first two training stages (1a - 2b), and all passed the subsequent preference test (3) required to move on to testing. However, two out of four birds reached extinction on opening the boxes before the end of testing (see Table 4.1). The prediction for this experiment was that, similarly to the food sharing experiment, the males would share a greater proportion of waxworms when the female was pre-fed mealworms than when she was pre-fed waxworms. However, as a full set of data was only available for two of the four birds, and these two birds showed opposite patterns, no conclusions can be drawn from these results (see Table 4.2).

There are two potential factors that may have contributed to the extinction of the string pulling behaviour, (i) the males were not directly rewarded for pulling the strings during testing and (ii) the females showed a low rate of taking the food once the boxes were opened. In the test trials, only the far bowl was baited, which was delivered to the females' compartment, so the males' motivation to pull the strings likely declined over time. However, it was predicted that the males would be motivated to the pull the strings in order to provide food to the female. This assumption may have failed as none of the females consistently took the food from the opened box; one female never ate from the box during testing (Jerusalem, see Table 4.1). This was despite the fact that during familiarisation (Stage 4) all females consistently approached and ate from the box within 2 minutes of opening.

Table 4.1 Data for Experiment 2, ordered by testing day. Males are in dark grey, showing the number of trials in which they pulled a string to open a box on each day, out of a maximum of 10 trials. Two out of the four birds (Hoy and Lima) reached extinction on the trained behaviour and stopped pulling strings before the end of testing. Females are in light grey below their male partner, indicating the number of trials in which they took food from the box when it was opened by their partner. Dashes indicate testing days on which their partner did not open any boxes and thus the female was unable to take any food from them.

	Test Day					
	1	2	3	4	5	6
Caracas	9	10	8	3	9	5
Wellington	8	6	0	2	1	2
Hoy	6	5	0	0	0	0
Hunter	5	5	-	-	-	-
Lima	6	1	4	0	0	0
Jerusalem	0	0	0	-	-	-
Romero	8	7	1	2	1	6
Webb	4	2	0	0	0	0

Table 4.2 Proportion of the boxes opened by the male jays that contained waxworms in each condition, calculated using the formula $\frac{\text{Waxworms}}{\text{Waxworms} + \text{Mealworms}}$. BL indicates the baseline condition, M indicates the condition in which females were pre-fed mealworms, and W indicates the condition in which the females were pre-fed waxworms. No value is shown for conditions in which the males failed to open any boxes across the 20 trials as this would require division by zero. If males were responding to the females' satiety they should share a greater proportion of waxworms in the M condition than the W condition.

	Condition		
	BL	M	W
Caracas	0.53	0.43	0.73
Hoy	0.55	-	-
Lima	0.43	0.25	-
Romero	0.47	0.67	0.57

A study on prosocial behaviour in jackdaws found that the behaviour of the potential recipient affected how likely an actor was to make a prosocial choice (Schwab et al., 2012). The actor could choose between opening a selfish box with a food reward only for the actor (1/0), or a co-operative box with a food reward for both the actor and the recipient (1/1). The jackdaws were more likely to choose the co-operative box when the conspecific positioned themselves next to the box to receive the food (Schwab et al., 2012). Even in an altruistic condition, the behaviour of the recipient could increase the likelihood of the actor choosing a reward for only the recipient (0/1) over a selfish reward (1/0), particularly if the recipient was of the opposite sex. Furthermore, in an experiment with chimpanzees, Melis et al. found that actors were much more likely to provide food to a recipient when the recipient actively signalled their interest in the food (Melis et al., 2011). Thus, it is possible that a lack of interest from the female might have led to a low motivation for the males to pull the strings in this experiment.

If the females do not eat the food and the males are not themselves rewarded, then there is no motivation for the males to pull the strings and thus it is unsurprising that they reached extinction on this behaviour. Romero opened a high number of boxes on the final testing day (6/10), despite his partner failing to take worms from the box. However, this could be explained by a trial on this day in which he pulled strongly on the string and the spinning arm detached. This allowed him to pull it through into his compartment and obtain the food from the far bowl. His motivation to interact with the boxes may have increased in subsequent trials as a result of this incident.

The fact that females tend not to take the food from the boxes is an interesting finding as it contrasts with their behaviour in food sharing experiments. In the context of food sharing the females willingly take both the sated and non-sated food from their partner, while in this experiment they showed limited willingness to take any food from the box apparatus. This may suggest that there is something specifically motivating about the food sharing behaviour itself, which encourages females to take food from their partner. In the next section I consider an experiment that could potentially explore this possibility.

4.4 Motivation Pilot

Studies of animal welfare have shown that animals have certain highly motivated behaviours, and that being prevented from performing these rewarding behaviours is a severe stressor (Boissy et al., 2007). Different animals will find different behaviours rewarding, for example chimpanzees have a greater intrinsic motivation to play with and manipulate objects than bonobos, which is likely a driver of differences in tool use between the two species (Koops,

Furuichi, & Hashimoto, 2015). In general, positive social interactions, such as maternal contact, social play and sexual interactions, have high incentive value and are a driving force of behaviour in both humans and non-human animals (Trezza, Campolongo, & Vanderschuren, 2011). It is possible that courtship behaviours, such as food sharing in Eurasian jays may be rewarding for individuals because of an association between effective courtship and reproductive success. In songbirds it has been shown that the production of sexually-motivated song is reinforced by the release of opioids during social interaction (Riters, 2012).

If the act of food sharing in Eurasian jays is itself inherently motivating regardless of which food is being shared, females will be likely to accept even undesirable food items from their partner. Thus, I have designed an experiment in which females' desire for a food type is manipulated by pre-feeding, followed by a chance to take either desirable or undesirable food from their partner (social condition) or from the platform in their compartment (non-social condition). If the females' inherent motivation to share counteracts the undesirability of the sated food, their preference for taking desirable food items should be lower in the social condition than the non-social condition. Here I describe a pilot study for this paradigm.

I conducted a pilot study first as it was necessary to determine whether the females would be willing to take worms in the non-social condition, and to assess the feasibility of the design. However, other researchers were conducting experiments with the Eurasian jay colonies, and so constraints on time and the availability of the birds meant that I did not have the opportunity to conduct a full study this breeding season. The pilot study revealed some potential alterations that could improve the experiment, which could potentially be conducted at some point in the future.

4.4.1 Subjects

Two male-female pairs from Colony 1 were tested in a pilot for this experiment in April 2019. The birds were housed in outdoor aviaries (20 × 6 × 3m) and tested in covered compartments (2 × 1 × 2m). They were fed a maintenance diet of soaked cat biscuits, vegetables, seeds and fruit and had ad libitum access to water. To ensure that the birds were mildly hungry and thus motivated to act for food, maintenance diet was removed from the aviary approximately 1 hour before the start of testing. Birds were tested once per day between the hours of 9am and 12pm, and the timing of testing for each pair was kept as similar as possible between days to keep their motivation consistent.

4.4.2 Methods & Analysis

During testing the male and female were in adjacent compartments with the mesh between them covered by transparent plastic sheeting. This was to prevent the birds from sharing anywhere except a window cut into the plastic at a specific location that was in clear view for the experimenter. All birds received a baseline trial in which the male was able to see the female eating maintenance diet. This baseline ensured that the pairs were motivated to share with each other in the experimental set-up. The passing criteria required the male to take a minimum of 10 worms from the experimenter (50%) and share with the female twice during the sharing phase. Only pairs that passed this baseline proceeded to the test trials.

Each of the two conditions (social and non-social) consisted of two trials, for a total of four test trials counterbalanced across four days. Each test trial consisted of a pre-feeding phase followed by a sharing phase run consecutively on the same day. The pre-feeding phase was identical across all four trials. A bowl containing 50 waxworms was placed in the females' compartment for 15 minutes, with the expectation that she would develop specific satiety to the waxworms and thus find them less desirable. The sharing phase differed between the two trials in each condition, involving the presentation of either 20 waxworms, the same as the pre-fed food (undesirable trial) or 20 mealworms, contrasting with the pre-fed food (desirable trial). The male was not present during pre-feeding, remaining in the outside aviary while the female was in the testing compartment. This should prevent the male from altering his food sharing based on the female's pre-feeding; ensuring that any difference in the amount of food taken by the female is due to her motivation rather than differences in what the male offers to her.

In the sharing phase of both the social and non-social conditions, the male was present in the compartment adjacent to the female. However, the manner in which the worms were presented differed between the two conditions. In the social condition, the procedure was similar to previous food sharing experiments. In each presentation one worm was held up to the mesh of the male's compartment for 30 seconds for the male to take, if the worm was not taken in 30 seconds it was removed. The presentation hand (left or right) was randomised to minimise the potential for hand biases forming that may interfere with any future food sharing experiments. This was followed by a 40 second inter-presentation interval, in which males could either eat, cache, or share the food with the female through the mesh window. This interval was maintained even when the male did not take the worm.

In the non-social condition, the worms were placed directly on the platform of the female's compartment so that the female could retrieve them. Similarly to the social conditions, the worm was left on the platform for 30 seconds and if the worm was not taken within 30 seconds

it was removed. This was followed by a 40 second inter-presentation interval to keep consistency between the rate of presentation in the social and non-social conditions. The worms taken by the female (either from the male or from the platform) were live-coded by the experimenter. The sharing phase for each condition was also videoed as a record of the experiment.

It was expected that there would be differences in the overall amount of worms the female would take between the non-social and social conditions. Therefore, rather than comparing the absolute number of undesirable worms taken by the female between the two conditions, I compared the number of worms accepted in the desirable trial relative to the total number of worms taken across both trials[14]. If the females find food sharing inherently motivating even when the shared food is undesirable, the females should show a greater preference for taking desirable worms in the non-social condition than the social condition. These values were considered descriptively, as this pilot only involved two male-female pairs.

4.4.3 Preliminary Results & Discussion

The results of the pilot study were inconclusive with regards to the effectiveness of this paradigm. Out of the two pairs tested, results could only be collected from one pair (Hoy and Hunter) as the other female (Webb) was not bold enough to take food from the platform in the presence of a human experimenter. Hunter's results show that she preferred to take the desirable food in both the social condition and the non-social condition (see Table 5.1). She did not show a smaller preference for the desirable food in the social condition than the non-social condition, which does not provide evidence in favour of the hypothesis tested here. However, it is impossible to draw any reliable conclusions based on a sample of one bird.

The fact that Webb took food from the male in the social condition but not from the platform in the non-social condition might suggest that the act of food sharing was sufficiently motivating to overcome the female's fear of approaching the experimenter. However, landing on the platform may have been more intimidating than landing on the food sharing perch as it required coming closer to the experimenter. An improvement upon this pilot might be to present the food in the non-social condition in the same location that food sharing occurs in the social condition, potentially through an automatic dispenser.

[14] This was calculated using the formula $\frac{M}{M+W}$, where M is the number of mealworms taken by the female (desirable trial) and W is the number of waxworms taken by the female (undesirable trial).

Another difficulty with this paradigm was the difference in the total number of worms taken by the female in the social and non-social conditions. In the non-social condition the female was consistently offered 20 worms, whereas in the social condition the number of worms the female could take varied depending on the male's food sharing. The male was given 20 worms, however in all food sharing experiments he consistently shares fewer with the female, between 0 and 15 in published studies with a mean of 4.93 out of 20 (Ostojić, Legg, et al., 2016; Ostojić, Legg, Shaw, et al., 2014; Ostojić et al., 2013). I accounted for this by comparing the proportion of desirable worms taken by the females between the conditions rather than the absolute difference. However, a closer comparison might be achieved by repeating the social condition and summing the food taken across repeats. Alternatively, a number of food sharing trials could be conducted with that pair in advance of the test trials, allowing an average number of sharing events to be calculated. The number of presentations in the non-social condition could then be matched to the approximate rate of sharing for that male.

Finally, it is also possible that the male and female may have different motivation to engage in food sharing behaviour. This experiment was designed only to test the female's motivation, but it is possible that the male may place a greater value on food sharing than females, as courtship behaviour is usually driven by males attempting to secure female partners. Therefore, in addition to a full experiment assessing the females' motivation, it would also be beneficial to assess the males' motivation if an appropriate experiment can be designed.

Table 5.1 Results for the pilot study, M indicates the condition in which the females were able to take mealworms (from the male or the platform) and W the condition in which the females were able to take waxworms. Prop indicates the proportion of desirable worms (mealworms) taken by the female in each condition. The female was always pre-fed waxworms, so mealworms should always be the desirable worm type. No value is shown for non-social condition for Webb as she did not take any worms from the platform. If food sharing was inherently motivating for the female regardless of the type of food being shared the proportion should be smaller for the social condition than the non-social condition.

		Social			Non-social		
Female	Male	W	M	Prop	W	M	Prop
Webb	Romero	2	7	0.78	14	20	0.59
Hunter	Hoy	0	2	1.00	0	0	-

4.5 General Discussion

In this chapter I investigated paradigms that could be used to explore Eurasian jays' sensitivity to their partners' specific satiety outside of the context of food sharing. However, the paradigms trialled here reveal a limitation of such novel contexts. In Experiment 1, I found that jays were not sufficiently motivated to look through peepholes in anticipation of their partner to allow for a paradigm involving looking measures. Furthermore, in Experiment 2 I found that in a food provision paradigm females were not motivated to take food, potentially because of the pre-feeding. This contrasts with beak-to-beak food sharing where females will take food from their partner even after pre-feeding. I therefore piloted a paradigm to assess the inherent motivation underlying food sharing behaviour, regardless of the desirability of the shared food. This pilot may provide the basis for an improved version of this experiment that could be employed by other researchers in the future.

In order to assess whether Eurasian jays' co-operative desire sensitivity is specialised or generalised, this ability needs to be explored in novel contexts. However, this can lead to problems with attention and motivation. Chimpanzee researchers have emphasised the necessity of testing cognitive abilities related to theory of mind in 'ecologically valid' scenarios; specifically referencing that chimpanzees may perform better on competitive than co-operative tasks involving mind-reading (Hare, 2001; Hare & Tomasello, 2004). However, others have argued that this assumption that chimpanzees are more likely to demonstrated their cognitive skills in competitive situations than co-operative situations is unfounded (Penn & Povinelli, 2007). In addition, as described in Chapter 1, there is a significant issue with mind-reading experiments that rely on familiar situations. As Povinelli and Vonk put it, "behavioural interactions that make the most ecological sense to the organism are precisely the ones that will be least diagnostic of whether the organism is reasoning about mental states" (pg. 159)(Povinelli & Vonk, 2003). If the individual being tested has had specific experience with similar scenarios in the past it is highly likely they will have had the opportunity to learn behavioural rules that they can use to directly predict how others will behave. Furthermore, 'ecologically valid' contexts are those that are most likely to have evolved highly specialised behaviours, which might allow adaptive responses without any representation of internal mental states.

Experiments proposed to solve this problem, such as Povinelli's 'bucket test' (see Chapter 1), have been described as 'unnatural' and thus unlikely to be a fair test of chimpanzees abilities (Hare, 2001). Additionally, other researchers have raised concerns that the increasingly complicated paradigms that are being proposed to distinguish mind-reading and behaviour-reading are based on setups that are known to be difficult for chimpanzees to understand (van

der Vaart & Hemelrijk, 2012). For example, the bucket test relies on an understanding that a bucket over the head prevents individuals from seeing, however existing evidence already suggests that chimpanzees are unable to understand this scenario (Povinelli & Eddy, 1996; Reaux et al., 1999).

The results of the experiments presented in this chapter highlight the difficultly in balancing ecological validity with designing effective experiments that might be able to assess the cognitive abilities of non-human animals. In this chapter, I was unable to find a paradigm to test Eurasian jays' sensitivity to their partner's specific satiety in a non-food-sharing co-operative context. However, one area in which Eurasian jays' have already been shown to respond to others' desires is that of food caching (Ostojić et al., 2017). As it is more closely relates to jays' natural behaviours, this existing caching study may represent a more ecologically valid test of their social cognitive abilities than the paradigms attempted in this chapter. In Chapter 5, I therefore attempt to expand on this study, exploring further the possibility that Eurasian jays' may be sensitive to the specific satiety of non-partners in a competitive caching context.

5. Chapter 5 – Eurasian jays' attribution of specific satiety in a competitive context

5.1 Introduction

In Chapter 4 I focused on Eurasian jays' sensitivity to the desires of others in a co-operative context, expanding upon food sharing experiments that assess interactions between partners. However, there is also evidence that Eurasian jays employ social cognitive abilities in a competitive caching context. A number of studies in jays have provided evidence for flexible cache protection strategies that may reduce the likelihood of caches being pilfered (Grodzinski & Clayton, 2010). It has been demonstrated that scrub jays act in ways that limit the visual information about the location of their caches available to observers: caching preferentially in a poorly lit location, in a far location (Dally, Emery, & Clayton, 2004b), or behind a barrier (Dally, Emery, & Clayton, 2005). In addition, scrub jays also appear to be more likely to re-cache their food in new locations after they have been observed during caching (Emery & Clayton, 2001).

Similarly, Eurasian jays seem to act in ways that might reduce both the visual and auditory information available to observers: preferentially caching in a far location (Legg, Ostojić, & Clayton, 2016), behind a barrier (Legg & Clayton, 2014), or in a quieter substrate (Shaw & Clayton, 2013). Crucially, these studies find a difference in the birds' caching strategy when they cache in private versus in the presence of a conspecific observer. This indicates that these cache protection strategies are related to the presence of the observer, and not just due to a general, inflexible preference for caching in less exposed or quieter locations.

Of particular interest to my thesis is a recent study that tested both scrub jays and Eurasian jays; investigating whether a cacher may be sensitive to the specific satiety of an observer (Ostojić et al., 2017). It was suggested that an observer that is sated on a certain food may be less likely to pilfer that food. Thus, a successful strategy for a cacher may be to preferentially cache the food that an observer is sated on. Ostojić et al. first confirmed their assumption that a bird's pilfering was affected by their specific satiety. A pilferer was pre-fed on either macadamia nuts or peanuts and shown a human experimenter caching both types of nuts in a tray. The pilferer was then given the opportunity to pilfer from the tray. It was found that jays showed a greater preference for pilfering peanuts when pre-fed macadamia nuts than when pre-fed peanuts. In a subsequent caching experiment the authors then found that the jays showed a greater preference for caching peanuts when an observing conspecific had been pre-fed peanuts than when they had been pre-fed macadamia nuts.

This suggests that the cache protection strategies of scrub jays and Eurasian jays go beyond simply reducing available information about the location of caches. They also appear to alter their caching in response to an observer's desires. This is particularly interesting as the context is competitive rather than co-operative so the caching experiment involves the opposite response to previous food sharing studies (Ostojić et al., 2013). Rather than choosing to share the food that their partner desires, here the jay must *avoid* caching the food that the observer desires. This indicates that the jays' sensitivity to specific satiety may be a flexibly applied social ability that can be used in different contexts.

However, it is still unclear what cognitive mechanisms underlie these cache protection strategies. It has been suggested that such findings imply that the cacher has some understanding of the perspective of the observer, and thus cache protection strategies result from mental state attribution (Clayton et al., 2007; Grodzinski & Clayton, 2010). In contrast, others suggest that these strategies could be explained by responses to observable cues regarding the presence or behaviour of the observer during caching (Butterfill & Apperly, 2013; Penn & Povinelli, 2007).

As in their original food sharing study (Ostojić et al., 2013), in the caching study Ostojić et al. attempted to test for potential behaviour-reading explanations by including a seen condition and an unseen condition. In the seen condition the cacher was able to watch the observer eat the pre-feeding food, while in the unseen condition visual access was blocked between the cacher and observer during pre-feeding. This meant that in the unseen condition the cacher would not know what the observer had been pre-fed, and the only way they could react to the observer's satiety would be based on the observer's behaviour during caching. In contrast to the food sharing study (Ostojić et al., 2013), the caching study found that the cachers' caching differed in line with the observer's satiety in both the seen and unseen conditions, suggesting the observer's behaviour during caching was sufficient to alter the cacher's response. However, this does not rule out the possibility that they are also capable of responding to other cues.

Here I describe an experiment that follows from the original caching study, attempting to determine whether Eurasian jays are sensitive only to the observer's behaviour at the time of caching or also to which type of food the observer had been pre-fed (Experiment 1). I focused specifically on Eurasian jays, despite the original study using both scrub jays and Eurasian jays, as there is no longer a population of scrub jays at the University of Cambridge. The experimental design was based on an unpublished study asking a similar question in the food sharing context (Ostojić, Legg, Mendl, et al., 2014). In addition, in Experiment 2, I describe an attempt to replicate the original findings from Ostojić et al.'s (2017) caching study, repeating both the pilfering experiment (2a) and the caching experiment (2b).

5.2 Experiment 1

In the seen condition of the original caching experiment (Ostojić et al., 2017), the cacher could potentially have been responding to observable behavioural cues from the observer rather than representing their desire. There are two possible points at which the cacher had the opportunity to detect behavioural differences: (i) changes in eating during the pre-feeding phase such as a decreased rate of eating or rejection of the remaining food, and (ii) changes in attention during caching due to differences in interest in the two foods. In the unseen condition of the original study the pre-fed food was unknown to the cacher and they could not watch the pre-feeding, preventing them from using any pre-feeding behavioural cues (i). However, jays still showed a caching response in line with the observer's satiety, thus, this suggests that the jays were able to respond to the observer's satiety based only on the behavioural cues expressed by the observer during caching (ii).

This finding appears to favour a behaviour-reading explanation. However, while the behavioural cues from the time of caching may be sufficient to produce a difference in the cacher's caching pattern, this does not mean that this is the only type of cue that the jays are able to respond to. It is possible that Eurasian jays may also be sensitive to the type of food observers are pre-fed, in addition to responding to their behaviour at the time of caching. In order to test this possibility, I therefore employed a paradigm based on an unpublished food sharing study (Ostojić, Legg, Mendl, et al., 2014), in which information about which food the observer was pre-fed was in opposition to the behavioural cues produced by the observer.

If visual access is blocked between the cacher and the observer's compartments during pre-feeding, a bowl can be shown entering the observer's compartment and then, after a delay, coming out empty. From this sequence of events the cacher could potentially infer that the food was eaten by the observer. However, the actual pre-feeding food could be switched to a different food out of view of the cacher. The observer's behaviour during caching would then indicate the opposite satiety to that suggested by the type of food that the cacher saw entering the compartment. There are three possible outcomes from this experimental design:

1. The cachers only respond to the behavioural cues produced by the observer during caching: the cachers' caching pattern would be dependent only on the food that the observer was actually pre-fed in both the matching and conflicting conditions, regardless of which food the cacher saw entering the observer's compartment during pre-feeding.
2. The cachers prioritise the available information about the pre-feeding: the cachers' caching pattern would be driven primarily by which food the cacher saw entering the

observer's compartment during pre-feeding, regardless of which food the observer had actually been pre-fed.

3. The cachers respond to both the behavioural cues and the available information about the pre-feeding: the cachers' caching pattern would be in line with the observer's actual satiety in the matching condition, but not necessarily in the conflicting condition as the cacher would also take into account which food they saw entering the observer's compartment.

5.2.1 Subjects

Five birds from two colonies acted as cachers in this experiment (Colony 1 = 4, Colony 2 = 1), with one further bird from Colony 2 acting only as an observer. The experiment was conducted from January to February 2018. Birds were housed in outdoor aviaries (20 × 6 × 3m) and tested in compartments (2 × 1 × 2m). Birds were fed a maintenance diet of soaked cat biscuits, vegetables, seeds and fruit and had ad libitum access to water. To ensure that the birds were mildly hungry and thus motivated to act for food, maintenance diet was removed from the aviary approximately 1 hour before the start of testing. The birds were tested between the hours of 9am and 12pm, with a follow-up session in the afternoon between 1 and 2pm. The timing of testing for each pair was kept as similar as possible between days to keep their motivation consistent. The birds from Colony 2 were tested by Ljerka Ostojić.

5.2.2 Methods & Analysis

During testing the observer and cacher were in adjacent compartments. Visual access between the two compartments was blocked by opaque plastic sheeting except for a cut out viewing window. Each test trial consisted of a pre-feeding phase followed by a caching phase, run consecutively on the same day. In the pre-feeding phase, a bowl containing either 50 macadamia nuts or 50 peanuts was placed in the observer's compartment for 15 minutes, with the expectation that they would develop specific satiety to that type of nut. Macadamia nuts were always given to the birds as quarters and peanuts as halves.

The information available to the cacher about which type of nuts the observer was given was manipulated between the two conditions. In both conditions the cacher was shown a bowl of 50 nuts in front of their compartment, and the bird's attention was called by the experimenter saying "Look, look [bird's name]". This bowl was then put into the observer's compartment, and the bowl was shown to the cacher through the viewing window before being placed on the platform. A second bowl containing 50 of the other type of nuts was shown to the cacher and

placed outside of the observer's compartment so that the cacher always saw both types of nuts in the pre-feeding phase. This should ensure that any response by the cacher would result from the food being placed in the observer's compartment rather than simply the sight of the food.

An opaque barrier was then put over the viewing window preventing the cacher from seeing into the observer's compartment. The food in the bowl was quietly removed into an empty bag and replaced with 50 nuts taken from a different bag kept in the experimenter's sleeve. In the matching condition the nuts were replaced with the same type of nuts, while in the conflicting condition they were switched to the other type of nuts. The switch was performed in both conditions to ensure that the timing and auditory cues were similar in both scenarios. The bowl was also lined with white cloth to minimise the noise caused by moving the nuts in case the birds were able to use sound to distinguish between the types of nuts.

Once the nuts had been successfully swapped, a bowl of maintenance diet was placed in the cacher's compartment and the birds were left for 15 minutes of pre-feeding. At the end of the 15 minutes any food remaining in the observer's bowl was moved into a bag hidden in the experimenter's sleeve to prevent the cacher from seeing any remaining food. The opaque barrier was then removed from the viewing window so that the cacher could see into the observer's compartment. The empty bowl was shown to the cacher through the viewing window and then again from outside their compartment. The bowl of maintenance diet was also removed from the cacher's compartment.

In the subsequent caching phase, the cacher was given the opportunity to cache while being watched by the observer. A 3x5 caching tray was placed in the cacher's compartment with a bowl of 50 macadamia nuts and a bowl of 50 peanuts. The caching trays consisted of spray-painted seedling trays containing removable, sand-filled seedling pots for the birds to cache in. The colour of the caching trays was pseudo-randomised such that the birds experienced a different coloured tray in each trial. A perch was placed on the mesh of the viewing window in the observer's compartment so that they could watch the cacher's actions. The birds were left for 15 minutes, to give the cacher an opportunity to cache both food types. At the end of the 15 minutes the experimenter returned and removed the caching tray. The experimenter released the jays into the main aviary and recorded the number, nut type and locations of any caches made by the cacher. The cachers were also videoed during the caching phase as a record of the experiment.

After testing, the compartments were left open so that all birds in the colony had access to them for at least an hour. This should allow any extraneous caches not made in the tray to be pilfered by other birds, a strategy included in all caching experiments to encourage cachers to cache in

the tray rather than in other locations around the compartment. Subsequently, the caching birds were given a chance to recover their caches from the caching tray, a second strategy to maintain the birds' motivation for caching in the trays. The cachers were shut into a run, separated from the other birds in the main aviary, and given access to the testing compartment through an open hatch. Their caching tray was placed in the compartment and the cacher was left for 10 minutes to give them the opportunity to recover their caches. At the end of the 10 minutes the experimenter returned and removed the caching tray. The experimenter released the jay into the main aviary and recorded the number, nut type and locations of any caches recovered by the cacher. The cachers were also videoed during recovery as a record of the experiment.

Each of the two conditions (matching and conflicting) consisted of two trials, one trial in which the observer was pre-fed macadamia nuts, and one in which the observer was pre-fed peanuts. This gave a total of four test trials occurring across four different days. Birds were not tested for more than two consecutive days without a day off, in order to maintain their motivation to cache. The order of conditions was counterbalanced across birds, and the order of trials within each condition was randomised. The trials were repeated on a subsequent day if the cacher failed to cache, or if the observer failed to eat the pre-feeding food. Each trial was to be conducted up to a pre-determined maximum of three times.

In the matching condition the behavioural indications of the observer's satiety should match with the information available to the cacher about which nuts the observer was pre-fed. In contrast, in the conflicting condition the behavioural indications of the observer's satiety should conflict with the information available to the cacher about which nuts the observer was pre-fed. Thus, the two conditions should produce different patterns of results depending on whether the cacher's response to the observer's specific satiety was based on the observer's behaviour or their information about the pre-feeding.

The data were analysed using paired permutation tests to determine whether there was a difference in the relative preference for caching peanuts (peanuts cached – macadamia nuts cached) between the trial in which the observer was pre-fed peanuts and the trial in which the observer was pre-fed macadamia nuts. Permutation tests were chosen to match the analysis used by Ostojić et al. (Ostojić et al., 2017). As in the original study, it was predicted that the jays would show a relative preference for caching the food the observer was sated on. However, it was unknown which cues about the observer's satiety the cacher would respond to: only the behavioural cues (outcome 1 above), primarily which food they saw entering the compartment (outcome 2) or a combination of both sources of information (outcome 3).

The matching condition was analysed using a one-tailed permutation test, as both cues were in agreement and it was a conceptual replication of the original finding so should follow the predicted pattern. In contrast, in the conflicting condition cachers the cues provided contradictory information and the difference could go in either direction so this was analysed with a two-tailed permutation test. This was then followed by a further paired one-tailed permutation on the difference of differences between the two conditions (matching and conflicting). A one tailed-test was used as the contradictory cues in the conflicting condition should make it more difficult for the cacher to respond to the observer's true satiety than in the matching condition, resulting in a greater difference between the two trials in the predicted direction in the matching condition than in the conflicting condition.

5.2.3 Results & Discussion

The jays did not show a greater preference for caching peanuts when the observer was pre-fed peanuts than when the observer was pre-fed macadamia nuts in either the matching condition (paired one-tailed permutation test, n = 5,T = 4, p = 0.750) or the conflicting condition (paired two-tailed permutation test, n = 5, T = 47, p = 0.562) (see Figure 5.1a). The difference in the preference for caching peanuts between trials was not greater in the matching condition than the conflicting condition (paired one-tailed permutation test, n = 5, T = 15, p = 0.719) (see Figure 5.1b).

These results were surprising as the matching condition was a conceptual replication of the unseen condition in the original caching study (Ostojić et al., 2017). If the jays are able to respond to the specific satiety of the observer based only on their behaviour during caching, as suggested by the original study, it was predicted they should show this caching preference in the matching condition. Furthermore, the matching condition of this experiment included an extra cue in comparison to the unseen condition of the original study, as they saw the bowl of nuts entering the observer's compartment and coming out empty, whereas in the unseen condition of the original study they never saw the food that the observer was pre-fed.

One possible explanation for this finding was inconsistent caching. Trials had to be repeated for two out of five birds because of low caching rates. For Purchas, three of the four trials had to be repeated (one of them twice) as she did not cache. For Romero, one trial had to be repeated twice, once because he did not cache and once because his observer (Hunter) failed to eat any of the pre-feeding food. A low rate of caching overall may have increased the likelihood of a false negative result. Furthermore, given the small population of available birds even if more individuals were tested it would be impossible to reach a significant effect with these results.

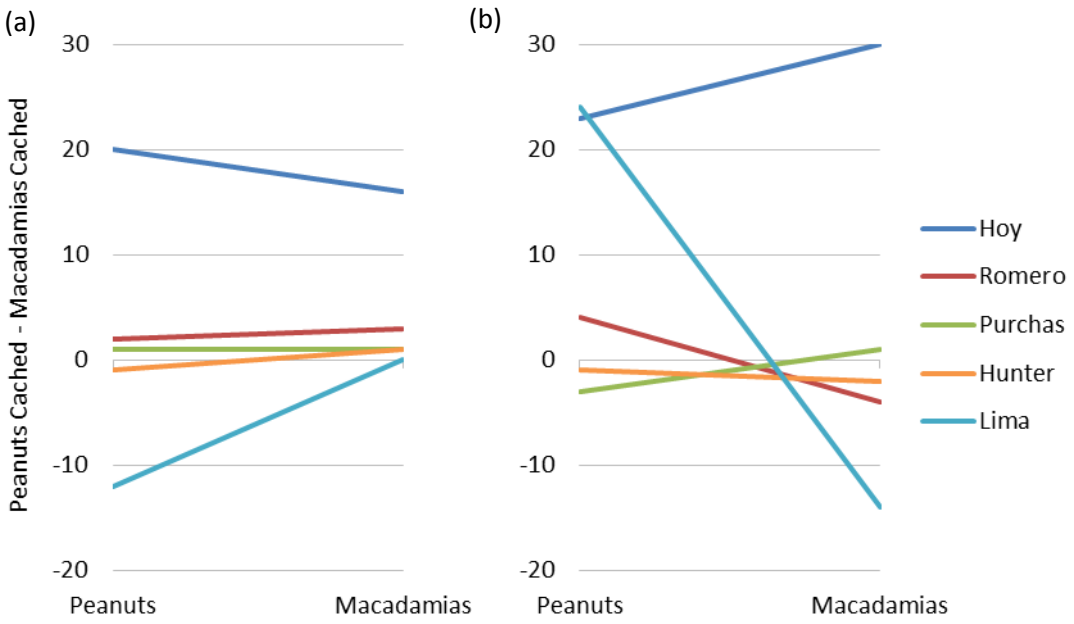

Figure 5.1 Graphs showing the difference between the number of peanuts cached and the number of macadamia nuts cached in each of the two trials: pre-fed peanuts and pre-fed macadamia nuts, (a) shows the results of the matching condition and (b) the results of the conflicting condition. If the jays were responding to the actual specific satiety of the observer the preference for caching peanuts should be higher in the pre-fed peanuts trial than the pre-fed macadamia nuts trial.

Alternatively, the finding from the unseen condition in the original study may have been a false positive for Eurasian jays due to the pooling of the analysis across both species. In the original study, it was found that scrub jays and Eurasian jays altered their caching strategy both when they could watch the observer eat the pre-fed food (seen condition) and when they could not (unseen condition). Therefore, it was expected that the Eurasian jays would be able to alter their caching strategy in the matching condition of this experiment. However, if the Eurasian jays are analysed separately from the scrub jays in the original study the results are less clear. The scrub jays cached in line with the observers' specific satiety in both the seen and unseen condition, whereas the Eurasian jays cached in line with the observer's specific satiety in the seen condition but not the unseen condition, although there was no significant difference between the two conditions (Gelman & Stern, 2006). Given that the jays did not show the expected pattern in the matching condition of Experiment 1, I chose to replicate Ostojić et al.'s study to verify the original results.

5.3 Experiment 2a

Ostojić et al.'s original study involved two experiments; a pilfering experiment and a caching experiment. The pilfering experiment was designed to test the assumption that being sated on a food would decrease the jays' likelihood of pilfering that food. Interestingly, this validation step contrasts with other studies investigating cache protection strategies, as these never test whether such strategies are actually likely to reduce pilfering. For example, researchers describe caching at a distance or in a shady area as cache protection strategies (Dally, Emery, & Clayton, 2004a; Dally et al., 2005; Legg, Ostojić, & Clayton, 2016), but researchers have not actually measured whether observers are less effective at pilfering caches hidden in these locations. By testing the assumption that pre-fed observers are less likely to pilfer the sated food, Ostojić et al. ensured that their assumption was valid before assessing the cacher's caching pattern.

The original pilfering experiment involved 16 birds, 10 scrub jays (4 female, 6 male) and six Eurasian jays (3 female, 3 male). The data from the Eurasian jays, when analysed separately from the scrub jays, show that the Eurasian jays were more likely to pilfer peanuts when pre-fed macadamia nuts than when pre-fed peanuts (paired one-tailed permutation test, $T = 0$, $p = 0.0313$). This confirms for the caching experiment that an observer should prefer to pilfer the non-pre-fed food, and thus supports the assumption that cachers should benefit from caching a food that the observer (a potential pilferer) is sated on.

In Experiment 2a, I have attempted to replicate this finding in Eurasian jays to confirm that caching an observer's sated food would be a valid cache protection strategy. Given the small size of the captive population available, some of the birds used in this replication were the same birds as those used in the original study. Thus the replication was not fully independent, and the generalisability of the results may be limited. However, the replication should at least show whether the effect is repeatable in a similar sample.

5.3.1 Subjects

The final subjects for this experiment were seven birds from two colonies (Colony 1 = 4, Colony 2 = 3). A further seven birds were tested but not included in the analysis; two of these birds were unable to complete the experiment for welfare reasons (Webb and Rome), one was excluded as they did not pass the baseline (Quito) and the other four were excluded as they did not pilfer any food in either of the trials (Ayton, Caracas, Dublin, Lisbon). Four of the birds tested here were also used as subjects in the original study (Caracas, Dublin, Lima and Washington). The experiment was conducted from September to December 2018. Birds were housed in outdoor

aviaries (20 × 6 × 3m) and tested in indoor compartments (2 × 1 × 2m). Birds were fed a maintenance diet of soaked cat biscuits, vegetables, seeds and fruit and had ad libitum access to water. To ensure that the birds were mildly hungry, and thus motivated to act for food, maintenance diet was removed from the aviary before the start of testing. The birds were tested once per day between the hours of 8am and 12pm. The timing of testing for each bird was kept as similar as possible between days to keep their motivation consistent. The birds from Colony 2 were tested by Piero Amodio.

5.3.2 Methods & Analysis

Each test trial consisted of a pre-feeding phase followed by a pilfering phase, run consecutively on the same day. In the pre-feeding phase, the bird was brought into the testing compartment and given a bowl of food for 15 minutes, with the expectation that they would develop specific satiety to that food. In the baseline trials the bowl contained a handful of maintenance diet, while in the test trials the bowl contained either 50 macadamia nuts, or 50 peanuts. Macadamia nuts were always given to the birds as quarters and peanuts as halves. In the two test trials a bowl containing the other type of nuts was placed outside of the compartment to ensure that any pilfering difference was caused by eating the pre-fed food, rather than simply the sight or smell of the food.

The jays were given the pre-feeding foods in different bowls to facilitate the distinction of the different foods (neutral bowl for maintenance diet, white bowl with red paw prints for peanuts, black bowl with spots for macadamia nuts). The colour of the bowls was the same for each bird and was kept consistent between experiments to facilitate birds identifying which nuts were in each bowl. The bowls were not counterbalanced between birds as the association itself was not part of the test and consistency of method reduces the chance of experimenter error. After 15 minutes the experimenter removed the bowl and any remaining food.

At the start of the pilfering phase the experimenter then placed a 5x5 caching tray in the compartment adjacent to the jay, with two bowls; one containing 8 macadamia nuts and the other containing 8 peanuts (colours as described above). The colour of the caching trays was pseudo-randomised such that the birds experienced a different coloured tray in each trial. The experimenter then cached these nuts in the caching tray in a pre-determined sequence at a consistent rate. The order of caches was pseudo-randomised so that the same type of nut was not hidden more than twice in a row and no pot contained more than one nut. The nuts were held up for the jay to see before being cached to increase the likelihood of the jay attending to each caching event.

A small transparent door connecting the two compartments was then raised to allow the bird to pass through into the compartment containing the caching tray. Once the bird was in the compartment the door was shut and the experimenter left, giving the bird 15 minutes to pilfer the caches from the tray. At the end of the 15 minutes the experimenter returned and removed the caching tray. The experimenter than let the jay out into the main aviary and recorded the number, type and locations of any nuts remaining in the tray. Nuts were coded as pilfered if they were either moved to a new pot in the tray or removed from the tray entirely. Nuts were coded as not pilfered if they were found in the same pot that they had been cached in by the experimenter, even if they had been uncovered by the pilferer.

The positioning of the caching tray during the experimenter's caching differed slightly from the original study (Ostojić et al., 2017). In Ostojić et al.'s study the caching tray was placed on a platform attached to the door of the bird's compartment. However, these platforms were no longer available at the time of the replication. Furthermore, the original study only tested birds from colony 2, and differences in the design of the aviary between the two colonies would prevent similar platforms from being attached to the doors in colony 1. Thus, in order to keep a consistent protocol between the two colonies I chose to place the caching tray inside the adjacent compartment. This difference should not affect the results of the experiment as the birds should be the same distance from the caching tray and be able to see it just as clearly as in the original experiment. The viewing angle of the birds to the tray would be slightly different; however 15 minutes should be enough time for the birds to search the whole tray so, as in the original experiment, their pilfering pattern should be affected more by their motivation to take the nuts than by their memory of their locations.

The baseline trial, in which the pilferers were pre-fed maintenance diet, was conducted first. This ensured that the birds were motivated to pilfer caches made by a human experimenter. The passing criterion required the birds to pilfer at least one nut from the tray. Only birds that passed the baseline proceeded to the test trials, if a bird failed to pilfer a nut after three baseline trials the bird was excluded from testing. The order of the two test trials (pre-fed macadamia nuts and pre-fed peanuts) was counterbalanced across jays. It was expected that the jays should show a greater preference for pilfering peanuts when they were pre-fed macadamia nuts than when they were pre-fed peanuts. This was analysed in the same way as the data in the original study. A paired one-tailed permutation test was performed to assess whether the number of peanuts pilfered minus the number of macadamia nuts pilfered was greater in the pre-fed macadamia nuts trial than the pre-fed peanuts trial.

5.3.3 Results & Discussion

The jays did not show a greater preference for pilfering peanuts in the pre-fed macadamia nuts trial than the pre-fed peanuts trial (paired one-tailed permutation test, n = 7, T = 7, p-value = 0.875, Figure 5.2b). This is a surprising result as it was expected that a pre-fed jay would be less likely to pilfer the sated food. This indicates that either: (i) the replication was a false negative, or (ii) the original result was a false positive.

A number of factors inherent to the field of comparative cognition can make it difficult to independently replicate studies. Most significantly, when working with animals sample sizes are often limited by space, funding and welfare concerns; reducing the number of animals available to be worked with. Furthermore, long-lived species such as chimpanzees and corvids are often used in a number of studies, and where paradigms can be repeated often the only sample available includes subjects tested in the original study (as seen here). This problem is specifically relevant in corvids as there are often a small number of research groups, or in some cases a single group, studying that species. These sample size issues can make it very difficult to find effects that are present because of low power. Additionally, if effects *are* identified in small samples the effect size is likely to be an overestimate relative to the true effect (Gelman & Carlin, 2014). This means that replications are likely to be underpowered even where the sample size matches or even exceeds that of the original study. As such, it is possible that the sample size used in this replication was insufficient, leading to a false negative result (i).

It is also possible that both the replication and the original study were hindered by a low level of pilfering. If the birds do not pilfer very much this leads to a floor effect making it harder to detect a difference between the two trials, increasing the likelihood of a negative result even if the effect is real (i). In each trial 16 nuts were cached by the human experimenter (8 peanuts and 8 macadamia nuts). However, in the original study, across the trials the average number of nuts pilfered was just 0.6 nuts for the scrub jays and 1.8 nuts for the Eurasian jays. In this replication the mean number of nuts pilfered was 4.9, but four birds were excluded from the analysis as they failed to pilfer in either of the two trials (Lisbon, Dublin, Caracas, Ayton), two of whom were included in the original study.

However, it is also possible that this low level of pilfering may have led to a false positive in the original study (ii), three out of the six Eurasian jays failed to pilfer any nuts at all in the pre-fed peanuts trial (Dublin, Rome, Jerusalem, see Figure 5.2a). Two of these jays showed a preference for pilfering peanuts in the pre-fed macadamia nuts trial giving a difference between the trials in the expected direction. However, Ostojić et al. state that the Eurasian jays had an overall preference for peanuts over macadamia nuts. As such, it is possible that the difference between

the trials was due to regression towards the mean. It is impossible to tell whether the lack of pilfering in the pre-fed peanuts trial gives a true indication of their relative preference for the two types of nuts, or was due to an unrelated lack of motivation to pilfer on that day. Thus, the results from these two birds may have appeared to differ in the expected direction by chance, resulting in a false positive effect in the original study.

If the pilfering pattern of observers is not in fact affected by their satiety, this has consequences for the interpretation of the original study. The pilfering experiment tests a key assumption for the caching experiment. Although the assumption was validated in the original study, here I did not replicate that result. This casts doubt on the value of a satiety-based cache protection strategy. If a sated observer does not show a reduced likelihood of pilfering the sated food, cachers would not benefit from caching the sated food over the non-sated food.

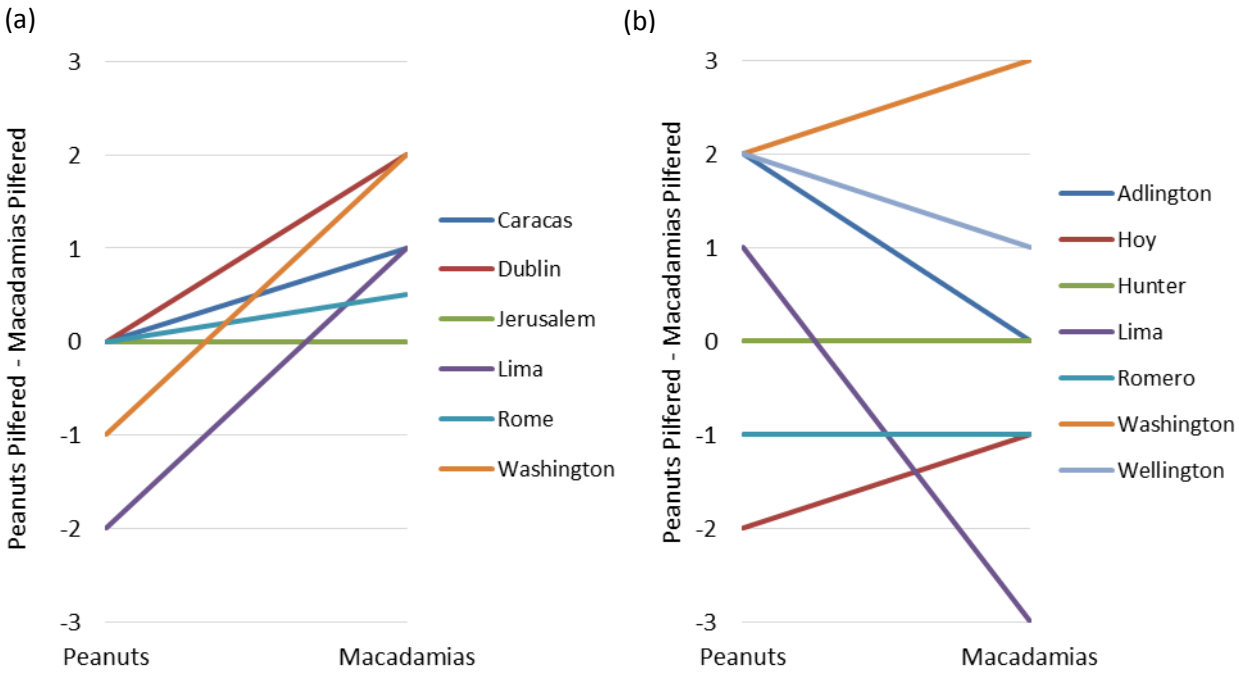

Figure 5.2 Graphs showing the difference between the number of peanuts pilfered and macadamia nuts pilfered in each of the two conditions: pre-fed peanuts and pre-fed macadamia nuts. (a) shows the results of the original study and (b) the results of the replication.

5.4 Experiment 2b

The second experiment in Ostojić et al.'s study was the caching experiment, which assessed whether jays would shift their caching strategy in line with the observer's specific satiety (Ostojić et al., 2017). The caching experiment involved 16 birds, 9 scrub jays (4 female, 5 male) and 7 Eurasian jays (2 female, 5 male). The study found that the pooled jays showed a greater

preference for caching peanuts when the observer had been pre-fed peanuts than when the observer had been pre-fed macadamia nuts in both the seen condition (one-tailed paired permutation test, n = 16, p = 0.006) and the unseen condition (one-tailed paired permutation test, n = 16, p = 0.003), and there was no difference between the conditions (one-tailed paired permutation test, n = 16, p = 0.255).

When the Eurasian jays are analysed separately they cached in line with the observer's specific satiety in the seen condition (one-tailed paired permutation test, n = 7, p = 0.047) but not in the unseen condition (one-tailed paired permutation test, n = 7, p = 0.109). However, there was no difference between the two conditions (one-tailed paired permutation test, n = 7, p = 0.242) (Gelman & Stern, 2006). Here, I attempted to replicate the caching experiment with Eurasian jays, in a sample at least as large as the original finding, to reassess whether they alter their caching in response to an observer's specific satiety.

5.4.1 Subjects

Subjects for this experiment were eight birds from two colonies (Colony 1 = 4, Colony 2 = 4). A further three birds were tested, two of these birds were excluded as they did not pass the baseline (Lisbon and Wellington), and one was excluded for failing to cache in both trials of one of the conditions (Rome). Of the birds involved in this study nine were also used in the original study. Two had been used as both cachers and observers (Hoy and Caracas), three just as cachers (Romero, Rome, Washington), and three just as observers (Adlington, Quito and Lima).

The experiment was conducted from October 2018 to February 2019. The birds were housed in outdoor aviaries (20 × 6 × 3m) and tested in compartments (2 × 1 × 2m). The birds were fed a maintenance diet of soaked cat biscuits, vegetables, seeds and fruit and had ad libitum access to water. To ensure that the birds were mildly hungry and thus motivated to act for food, maintenance diet was removed from the aviary before the start of testing. Birds were tested once per day between the hours of 8am and 12pm. The timing of testing for each bird was kept as similar as possible between days to keep their motivation consistent. The birds from Colony 2 were tested by Piero Amodio.

5.4.2 Methods & Analysis

During testing the observer and cacher were in adjacent compartments. Each test trial consisted of a pre-feeding phase followed by a caching phase, run consecutively on the same day. In the pre-feeding phase, the observer was given a bowl of food to eat from for 15 minutes, with the expectation that they would develop specific satiety to that food. In the baseline trials the bowl

contained a handful of maintenance diet, while in the test trials the bowl contained either 50 macadamia nuts, or 50 peanuts. Macadamia nuts were always given to the birds as quarters and peanuts as halves. In the two test trials a bowl containing the other type of nuts was placed outside of the compartment to ensure that any caching difference was caused by the observer's interaction with the pre-fed food, rather than simply the sight or smell of the food. The cacher was always given a bowl containing a handful of maintenance diet during pre-feeding. At the end of the 15 minutes the experimenter removed the bowls and any remaining food.

In the seen condition the cacher had visual access into the observer's compartment during pre-feeding. Thus, the cacher could see which food the observer was pre-fed and could watch them eating it. The pre-feeding food was presented in different coloured bowls in the same manner as Experiment 2a. In the unseen condition visual access was blocked by opaque white plastic sheeting in colony 1, or by an adjustable blue plastic blind in colony 2. Food was given to the jays in neutral coloured bowls, and the observer's bowl was covered until it was placed into the compartment, where it was uncovered to allow the observer to eat from it. The cacher was therefore unable to watch the observer eat the pre-feeding food and did not know which food the observer was pre-fed. The food outside of the compartment was also covered, so that the olfactory cues would remain the same but the food could not be seen.

At the end of the pre-feeding phase both jays were moved across one compartment through the connecting doors, using a waxworm to attract them. The observer was now in what was previously the cacher's compartment, which meant that they would not have access to any peanuts or macadamia nuts that they might have cached around the compartment. This should prevent the observer from being distracted by retrieving these nuts during the caching phase, and help prevent the cacher from seeing which type of nuts the observer had been pre-fed.

Once the birds were moved, the caching phase was initiated and the cacher was given the opportunity to cache in view of the observer. A caching tray was placed in the cacher's compartment with a bowl of 50 macadamia nuts and a bowl of 50 peanuts. The colour of the caching trays was pseudo-randomised such that the birds experienced a different coloured tray in each trial. In colony 1, 5x5 caching trays were used; whereas in colony 2, 3x5 trays were used. This difference resulted from a mistake in communication between myself and the experimenter. The birds were left for 15 minutes during which the cacher had the opportunity to cache both types of food. At the end of the 15 minutes the experimenter returned and removed the caching tray. The experimenter released the jays into the main aviary and recorded the number, nut type and locations of any caches made by the cacher. The cachers were also videoed during the caching phase as a record of the experiment.

On the following day, the caching birds were given a chance to recover their caches from the caching tray. The cachers were shut into a run, separated from the other birds in the main aviary, and with access to the testing compartment through an open hatch. Their caching tray was placed in the compartment and the cacher was left for 10 minutes to give them the opportunity to recover their caches. At the end of the 10 minutes the experimenter returned and removed the caching tray. The experimenter released the jay into the main aviary and recorded the number, nut type and locations of any nuts remaining in the tray. The cachers were also videoed during recovery as a record of the experiment.

The baseline trial, in which the jays were pre-fed maintenance diet, was conducted first. This ensured that the birds were motivated to cache in the experimental set-up. The passing criterion required the birds to cache at least one nut in the tray. Only birds that passed the baseline proceeded to the test trials, and the baseline was repeated up to a maximum of three times before the bird was excluded from testing. The order of the subsequent test trials (pre-fed macadamia nuts and pre-fed peanuts) was counterbalanced across jays. Additionally, the order in which the jays experienced the seen and unseen conditions was also counterbalanced across jays.

In the original study, Ostojić et al. state that they ordered the testing so that jays that were tested as both cachers and observers participated as cachers first (Ostojić et al., 2017). However, two birds (Pendleton and Hoy) acted as each other's observers meaning that at least one bird must have participated as an observer first. In this replication I chose to test the birds as cachers regardless of their previous experience of acting as observers, as otherwise the sample size would be too limited. A number of the birds available to me had already been used in the original study as observers (Quito, Lima, Adlington, Caracas and Hoy). Furthermore the number of birds in colony 2 was just seven at the time of the replication and some of the birds could not be used as observers as they show distress when shut in the testing compartments. At the start of testing it was planned that Hoy and Hunter would be tested with Romero as an observer; however this pairing had to be changed as neither bird will cache in the presence of Romero. This may be due to the position of Romero as the most dominant bird in the hierarchy of colony 2.

The results were analysed using paired one-tailed permutation tests to determine whether there was a difference in the relative preference for caching peanuts (peanuts cached – macadamia nuts cached) between trials in which the observer was pre-fed peanuts and trials in which the observer was pre-fed macadamia nuts. This was then followed up by a further paired one-tailed permutation test on the difference of differences in caching preference between the two

conditions (matching and conflicting). This matched the analysis used by Ostojić et al. in their original study (Ostojić et al., 2017).

If the jays are sensitive to the specific satiety of an observer they should show a greater preference for caching peanuts in the pre-fed peanuts trial than the pre-fed macadamia nuts trial. If the jays' caching was influenced by the observer's behaviour during the caching phase they should show the same caching pattern (caching relatively more of the sated food) in the seen condition and the unseen condition. This was the pattern identified in the original study. However, if the birds based their caching strategy on cues available at the time of pre-feeding they should show a difference in their caching pattern between the seen condition and the unseen condition, as in the seen condition they were able to watch the pre-feeding phase, while in the unseen condition their visual access was blocked.

5.4.3 Results & Discussion

The cachers did not show a greater preference for caching peanuts when the observer was pre-fed peanuts than when the observer was pre-fed macadamia nuts, in either the seen condition (one-tailed permutation test, $n = 8$, $T = 18$, $p = 0.570$, Figure 5.4a) or the unseen condition (one-tailed permutation test, $n = 8$, $T = 3$, $p = 0.981$, Figure 5.4b). The difference in the caching preference between the two trials was not greater in the seen condition than the unseen condition (one-tailed permutation test, $n = 8$, $T = 41$, $p = 0.156$, Figure 5.5). Thus, the results of the caching experiment did not replicate the results from Ostojić et al.'s original study. In fact, observationally in the unseen condition the cachers showed the opposite pattern with a greater preference for caching peanuts when the observer was pre-fed macadamia nuts than when the observer was pre-fed peanuts (see Figures 5.b and 5.5). However, as the analysis was one-tailed this was not directly tested and the only statistical inference that can be made is that there was no difference between the two trials.

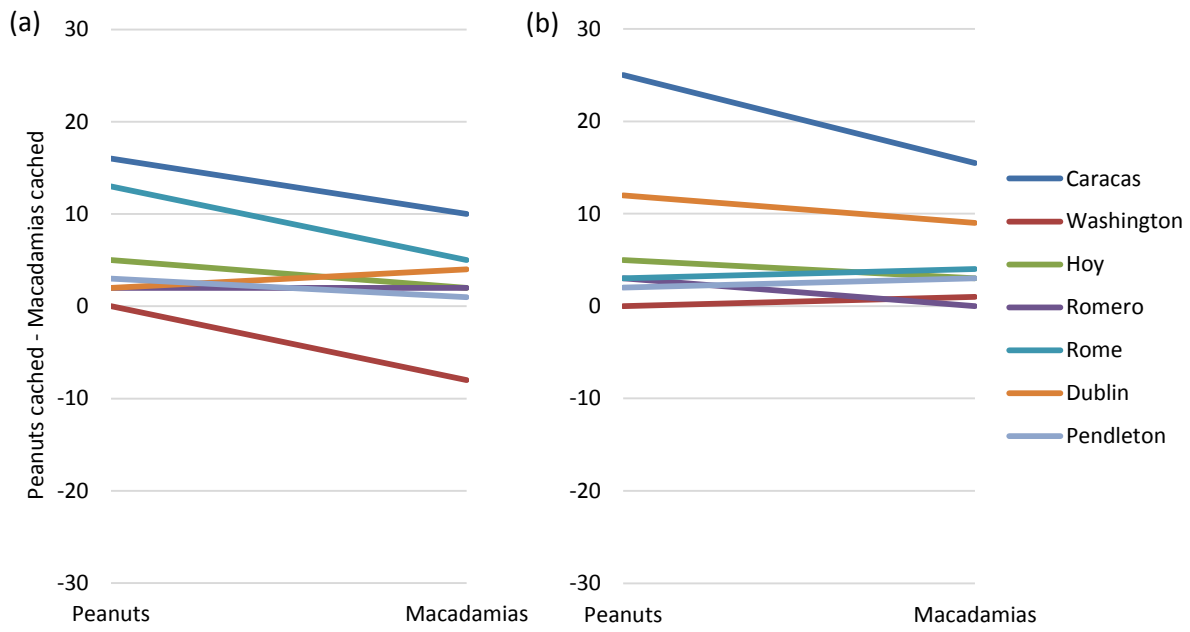

Figure 5.3 Graphs showing the difference between the number of peanuts and macadamia nuts cached in the two trials (pre-fed peanuts and pre-fed macadamia nuts) in the original study; (a) shows the seen condition and (b) shows the unseen condition.

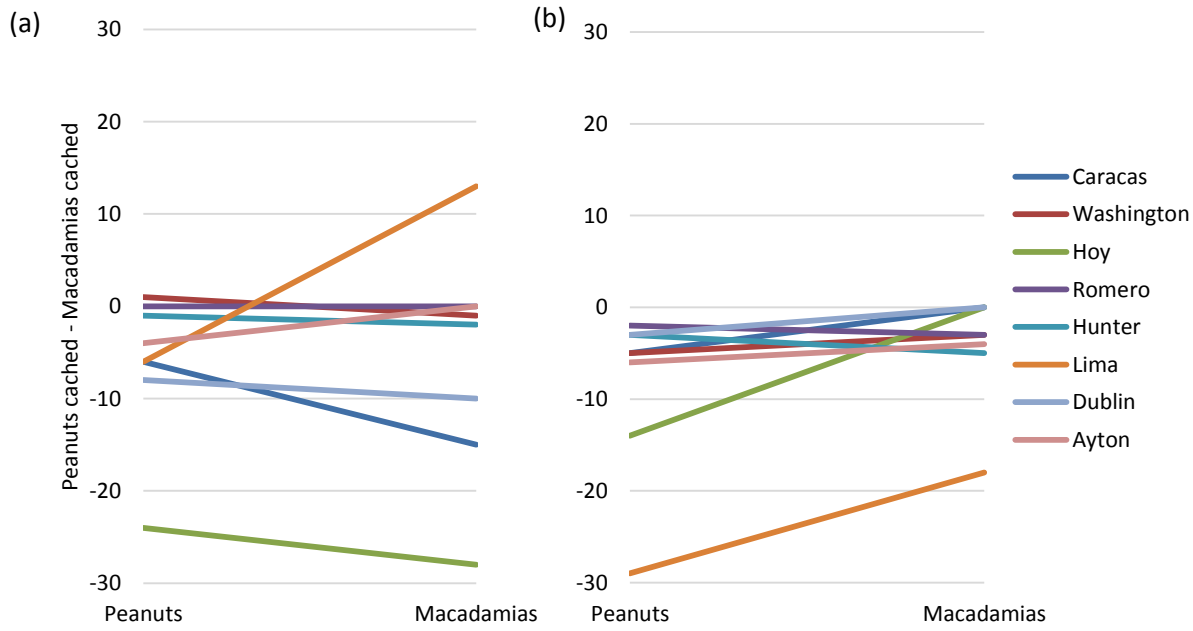

Figure 5.4 Graphs showing the difference between the number of peanuts and macadamia nuts cached in the two trials (pre-fed peanuts and pre-fed macadamia nuts) in my replication of the caching experiment; (a) shows the seen condition and (b) shows the unseen condition.

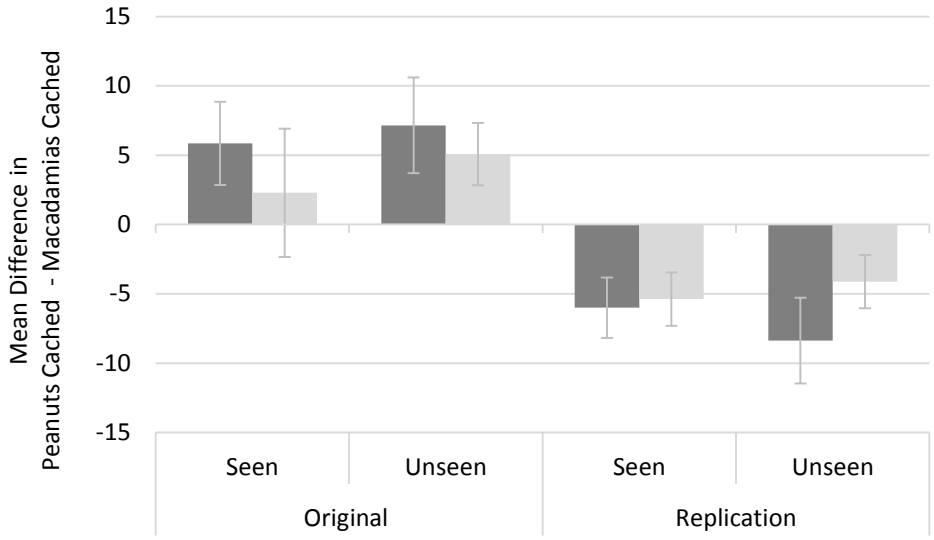

Figure 5.5 A graph showing the caching patterns of the Eurasian jays in each trial, from both the original caching experiment and the attempted replication (Experiment 2b). The bars represent the mean difference in nuts cached across jays (peanuts cached – macadamia nuts cached) in each condition. The dark grey bars show the trials in which the observer was pre-fed peanuts, and the light grey bars show trials in which the observer was pre-fed macadamia nuts. The error bars show the standard error of the mean.

In summary, I was unable to replicate either the pilfering experiment (2a) or the caching experiment (2b) from Ostojić et al.'s original study (Ostojić et al., 2017). The lack of an effect of the pilferer's satiety on pilfering preference in Experiment 2a calls into question the assumption that responding to the satiety of an observer would be a beneficial cache protection strategy. Combined with the lack of an effect on the caching response in Experiment 2b, this may suggest that cachers do not take into account the satiety of an observer. This therefore raises further questions regarding whether Eurasian jays have a generalised ability to respond to the desires of conspecifics or whether this is specific to their partner and/or the food sharing context (see Chapter 4).

5.5 General Discussion

In this chapter I attempted to expand upon a previous study in which Eurasian jays were shown to adapt their caching pattern to the specific satiety of a conspecific observer (Ostojić et al., 2017). The original finding indicated that the cachers' preference for caching different types of nuts was affected by the satiety of an observer even when the cacher had not seen the

observer's pre-feeding (unseen condition). This suggests that the behaviour of the observer at the time of caching was a sufficient cue to alter their caching preference. However, this does not preclude the possibility that they might also be able to respond to other cues about the desires of the observer. Therefore, in Experiment 1 I utilised a paradigm in which the behaviour of the observer during caching was manipulated such that it would match or conflict with the cacher's information about which food the observer had been pre-fed. It was expected that, as in Ostojić et al.'s study, jays would respond to the observer's satiety in the matching condition. This could then be compared with the conflicting condition. If the jays were sensitive only to the observer's behaviour they should respond the same in the conflicting and matching conditions, but if they were also sensitive to the information about which food the observer was pre-fed there would be a difference between the matching and conflicting conditions.

However, I found that the cacher did not cache in line with the observer's satiety in either the matching or the conflicting conditions. This was a surprising finding as the matching condition was a conceptual replication of the unseen condition in the original study. I therefore chose to replicate the original study in Eurasian jays, recreating both the pilfering experiment (2a) and the caching experiment (2b). In Experiment 2a I found that the jays did not show a decreased preference for pilfering a food they were sated on. Moreover, in Experiment 2b I found that jays did not show an increased preference for caching the food that a conspecific observer was sated on.

Ostojić et al.'s study provided the first evidence for Eurasian jays' sensitivity to specific satiety outside of the food sharing context. However, I was unable to replicate the results of the original study, which may lead to questions about the reliability of this effect. As described above, there are two possible explanations for the inability to find an effect here: (i) the replication was a false negative, or (ii) the original result was a false positive. It is difficult to draw a firm conclusion from these results as the small sample size limits the power of the study, and combined with high levels of noise it may be difficult to extract an effect even if it is present. The replication may have been underpowered as studies with small sample sizes can often overestimate the effect size (Gelman & Carlin, 2014). However, the effect found in the caching replication (Experiment 2b) was not just of a smaller size, but in the opposite direction (see Figure 5.5), suggesting that this factor may be insufficient to explain my negative results.

It is also possible that replication difficulties may be related to the issue raised in Chapter 4, i.e. Eurasian jays may be specifically motivated to respond to others' behaviour in a co-operative food sharing context. The low levels of pilfering and caching in both the original study and this replication may be due to a lack of motivation for the jays to produce the behaviours of interest.

As the birds are fed ad libitum throughout the year (outside of experiments) they may have a limited incentive to cache. Furthermore, they can cache outside of the caching trays in locations around the testing compartment, further reducing the caches that can be assessed. This contrasts with food sharing, which is likely to have a high incentive value and which, during testing, can only be conducted in a specific location through the window. As mentioned above a low level of pilfering or caching might make it harder to identify a difference between conditions, resulting in a false negative (i).

On the other hand, if the original result was a false positive (ii) then this implies that Eurasian jays do not respond to the specific satiety of an observer during caching. The original study involved a sample of both Eurasian jays and scrub jays, whereas this replication was only conducted with Eurasian jays. It therefore remains possible that scrub jays are capable of responding to the desires of others in a caching context, even if Eurasian jays are not.

Highly territorial species like Eurasian jays may be less likely to be observed by conspecifics while caching than a semi-territorial species such as scrub jays. As such there may be less evolutionary pressure for Eurasian jays to attend to others in the caching context (De Kort et al., 2012). However, such an observation remains speculative as there has been no systematic assessment of the differences in caching or pilfering behaviour between the two species. This is a relevant concern in comparative cognition research as replications may often be conducted in different species, meaning that any unexpected findings could potentially be attributed to species differences rather than a false positive/negative result (see Chapter 7).

If this result in Eurasian jays was in fact a false positive, this favours the possibility that Eurasian jays' sensitivity to others' specific satiety may be unique to the food sharing context, contradicting suggestions that jays have a generalised theory of mind system (Grodzinski & Clayton, 2010; Ostojić et al., 2017, 2013). Nonetheless, it is still possible that they might employ some form of desire attribution that is specific to beak-to-beak food sharing.

In summary, in this chapter I attempted to identify whether Eurasian jays' response to an observer's specific satiety in the caching context might involve sensitivity to cues other than the observer's behaviour at the time of caching. However, the jays did not respond as expected in the matching condition. In a subsequent attempt to replicate the original finding (Ostojić et al., 2017) I failed to recreate the predicted pattern. This failure to replicate may also reflect a wider problem with the reliability of cache protection strategy studies. It is possible that there is a 'file drawer effect' present in the field, with only positive results being published or disseminated (Rosenthal, 1979). In addition to the replication described here, I have also been involved in testing a second cache protection strategy in Eurasian jays, namely caching behind a barrier

(Legg & Clayton, 2014), which also failed to replicate (Amodio, 2019). It is likely that a number of these cache protection studies suffer from the same problems described above: small sample sizes and low rates of caching. Therefore, in Chapter 6 I chose to focus on the co-operative food sharing context in which male jays show a clear replicable response to their partner's behaviour, investigating whether Eurasian jays may be sensitive to their partner's choices between foods.

6. Chapter 6 – Sensitivity to the informativeness of others' choices

6.1 Introduction

In Chapter 5 I investigated Eurasian jays' social cognitive abilities in a competitive context. However, I failed to replicate previous findings regarding Eurasian jays' sensitivity to an observer's specific satiety when caching (Ostojić et al., 2017). I therefore return back to the co-operative food sharing context, in Chapter 6, expanding upon previous specific satiety studies. In previous chapters I have not found evidence for a sensitivity to others' desires outside the context of direct beak-to-beak food sharing. However, it is still possible that male jays' are able to respond flexibly to the desires and preferences of their partner as part of the food sharing response.

As I discussed in Chapter 1, the principle of goal-directed action in Apperly and Butterfill's minimal theory of mind system may account for infant's implicit responses to the preferences of others (Butterfill & Apperly, 2013, 2016). A number of studies utilising violation of expectation measures have suggested that infants may be sensitive to others' preferences (Luo, 2011; Luo & Baillargeon, 2005, 2007). Specifically, infants appear to show surprise when a protagonist stops choosing a previously preferred object (Luo & Baillargeon, 2005) (see also Chapter 2).

Furthermore, infants' response to others' preferences appears to go beyond simple statistical associations between actions and outcomes. For example, if A is chosen when B is present infants appear to expect the protagonist to subsequently choose A over B, whereas if A is chosen when B is not present or not visible they do not expect that A will be chosen over B (Luo, 2011; Luo & Baillargeon, 2007).[15] Such responses cannot be explained by simple associative processes as in both cases the protagonist is equally associated with A. The infant appears to be sensitive to the presence or absence of object B when the choices are being made; responding with 'surprise' only when the protagonist's previous choices were informative about which object was preferred over the other.

If this implicit sensitivity to the informativeness of others' choices involves a minimal theory of mind mechanism, it may also be present in human adults and non-human animals. As such, in this chapter I investigate whether Eurasian jays and human adult respond differently when they are shown a protagonist making uninformative and informative choices. If there are similarities

[15] There is conflicting evidence from another study which found that 9 month olds will anticipate an agent continuing to act towards the same goal when other options become available (Hernik & Southgate, 2012). However, criticisms have been made of Hernik and Southgate's conclusions, cf. (Luo & Choi, 2012).

between the response of infants, adults and jays to the informativeness of others' choices, this may provide evidence for a shared minimal system underlying these responses.

6.2 Investigating sensitivity to informativeness of choices in Eurasian jays

In previous food sharing studies involving Eurasian jays, males have been shown to alter their pattern of food sharing depending on what their partner had been pre-fed (Ostojić, Legg, et al., 2016; Ostojić, Legg, Shaw, et al., 2014; Ostojić et al., 2013). These studies all focused on manipulating the female's desires using the specific satiety phenomenon, in which the continued consumption of a single food reduces the desirability of that food (Balleine & Dickinson, 1998; Havermans et al., 2009). In order to share in line with the female's satiety males must reduce how much of the sated food he gives to the female compared to when she is sated on a different food type. As such, the food sharing pattern shown in experiments to date could theoretically result from a rigid rule where males are less likely to share food items that they have seen their partner eat recently. Although no attempt has yet been made to measure Eurasian jays' food sharing in the wild, the food sharing behaviour appears to be involved in courtship (Goodwin, 1951, 1956; Lack, 1940), and it is therefore likely to have a positive effect on reproductive success. Food sharing may therefore be under strong evolutionary pressure, allowing fixed innate responses to females' desires to arise over successive generations.

In this chapter I consider this possibility by assessing the flexibility of the food sharing response in male Eurasian jays. One unpublished study has already attempted to test whether the males' food sharing response is a flexible strategy (Legg, 2014b). This study was based on the assumption that when a female eats a small amount of a food it is unlikely that she is sated on that food. Instead, choosing to eat that food may indicate a preference for that food type. Thus, if male jays are able to flexibly respond to their partner's eating behaviour they might respond differently to their partner's choice based on the quantity of food eaten by their partner.

In Legg's experiment the female was given the opportunity to eat different quantities of waxworms and mealworms in a choice paradigm (Legg, 2014b). The female was presented with either a choice between a single waxworm and a single mealworm (1v1 condition) or a choice between 5 waxworms and 5 mealworms (5v5 condition)[16]. The choices of the female were controlled by coating one type of worm in quinine (an aversive substance), so that she would

[16] Five was chosen for the larger amount of worms as, while there was high variation in how many worms each female would eat, pre-tests showed that all females would eat at least five items of both worm types. This ensured that the bowls would be empty at the end of the choice phase in both conditions (1v1 and 5v5) so that the cues available to the male at this time would be the same across the different conditions.

only eat the worm(s) of the other type. In the 1v1 condition, the choice of one worm should indicate a preference for that type of worm over the other, and thus the male might be expected to share more of the chosen food. However, in the 5v5 condition, the choice and consumption of 5 worms may be more ambiguous as this number of worms might lead to specific satiety.

There was no difference in the proportion of waxworms shared between the two test trials (mealworm-eaten and waxworm-eaten) in either the 1v1 condition (two-tailed Wilcoxon signed rank test, n = 8, V = 20, p = 0.0625) or the 5v5 condition (two-tailed Wilcoxon, n = 8, V = 8, p = 0.999). However, a comparison of the 1v1 condition and the 5v5 condition found a difference in the proportion of waxworms shared across the two trials between the conditions (two-tailed Wilcoxon, V= 26, p = 0.0469). This was because, in the 1v1 condition, the males showed a trend to share a greater proportion of waxworms with the female when she chose a waxworm than when she chose a mealworm, while no such trend was seen in the 5v5 condition. This difference between the two conditions suggests that males' food sharing response may be flexible and dependent on the quantity of worms eaten by the female.

However, although the females were presented with a choice between two different food types (mealworms and waxworms), this experiment does not directly provide evidence that the males were responding to the female's choices. In the 1v1 condition it was predicted that the female's choice should indicate a preference for the chosen food, and thus the males' sharing pattern should shift in favour of the chosen food. However, in the 1v1 condition males did not share a statistically greater proportion of waxworms when a waxworm had been chosen than when a mealworm had been chosen.

Furthermore, the difference in the food sharing pattern between the 1v1 and 5v5 condition could be based solely on differences in the female's behaviour towards the chosen worms. As the female becomes sated the worms become less desirable and it is likely that her behaviour when taking the worms differs, e.g. the rate of her eating slows down. Therefore, there may be an observable difference in the female's behaviour when taking the chosen worm in the 1v1 condition and the 5v5 condition. It is possible that the female may respond more positively to the chosen worm(s) in the 1v1 condition than the 5v5 condition, causing a difference in the male's food sharing pattern without any need to take into account the choice of one worm type of the other.

Overall, while this experiment does suggest that the food sharing response of male jays is flexible to some extent; it does not clearly indicate that the male is responding to the choices of his female partner. I therefore present here an experiment to test whether the choice itself

might be relevant to the male's food sharing response, or whether the male is sensitive only to changes in the females' eating behaviour towards a single food.

6.3 Experiment 1

In Experiment 1, I investigated whether male Eurasian jays are able to respond to their partner's preferences, and in addition whether they are sensitive to the informativeness of their partner's choices. This paradigm was based on a previously pre-registered experiment designed by Brecht et al. (Brecht et al., 2016). In the informative condition the female was presented with choices between two items of different foods (A vs B), whereas in the uninformative condition the female was presented with a choice between two items of the same food (e.g. A vs A). Choosing A in the informative conditions and choosing A in the uninformative conditions should lead to equivalent association between the female and A. Thus, any differences in sharing pattern between informative and uninformative choices could not be explained by the interaction between the female and the chosen food. In order for there to be a difference between the conditions the male must additionally take into account the identity of the second worm (A or B). If the male is sensitive to the informativeness of their partner's choices it is predicted that they would show a greater preference for sharing the chosen worm in the informative condition than the uninformative condition.

6.3.1 Subjects

Six male-female pairs from two colonies (colony 1= 2 pairs, colony 2 = 4 pairs) were tested in April and May 2018, during the breeding season in which food sharing takes place with the highest frequency. All pairs had participated in previous food sharing experiments. There were two fewer pairs than in Legg's (2014) experiment as three birds from these pairs died of natural causes in the intervening period (Wilson – June 2014; Ohorougu – July 2014, Pendleton - January 2017). The birds were housed in outdoor aviaries (20 × 6 × 3m) and tested in indoor compartments (2 × 1 × 2m). They were fed a maintenance diet of soaked cat biscuits, vegetables, seeds and fruit and had ad libitum access to water. The birds were tested once per day. To ensure that the birds were mildly hungry and thus motivated to act for food, maintenance diet was removed approximately 1 hour before the start of testing. Colony 1 was tested between the hours of 10am and 1pm and colony 2 between 9am and 12pm. The timing of testing for each pair was kept as similar as possible between days to keep their motivation consistent.

6.3.2 Methods & Analysis

During testing the male and female were in adjacent compartments with the mesh between them covered by transparent plastic sheeting. This prevented the birds from sharing anywhere except a window cut into the plastic at a specific location that was in clear view for the experimenter. All birds received a baseline trial in which the male was able to see the female eating maintenance diet. A bowl containing maintenance diet was placed in the female's compartment and then removed as soon as the female had eaten from the bowl. This baseline ensured that the pairs were motivated to share with each other in the experimental set-up. The passing criteria required the male to take a minimum of 10 worms from the experimenter (50%) and share with the female at least twice during the sharing phase. Only pairs that passed this baseline proceeded to the test trials.

Each of the test trials consisted of a choice phase followed by a sharing phase, run consecutively on the same day. In the choice phase, the female was presented with two worms placed directly on the wooden platform in the testing compartment. The choices presented to the female were either informative or uninformative. In the informative choice condition the female was presented with one waxworm and one mealworm, while in the uninformative choice condition the female was presented with either two mealworms or two waxworms. In total each pair experienced four test trials: the informative condition with one waxworm chosen (W vs M) and one mealworm chosen (M vs W), and the uninformative condition with one waxworm chosen (W vs W) and one mealworm chosen (M vs M). Once the female had taken one worm, the experimenter removed the other worm to prevent the female from eating both worms. The worms were shown clearly to the male before the female made her choice, the male could watch her while making the choice, and the remaining worm was shown to the male after her choice to ensure that he saw which worm she had chosen. After the female's choice, perches were placed onto the mesh window in both compartments to facilitate sharing.

A version of this experimental procedure was pre-registered on the Open Science Framework (Brecht et al., 2016). However, the method used to manipulate the female choices was changed. The pre-registration suggested coating one type of worm in quinine, but instead I chose to exploit the birds' natural side bias for picking the worm to their right rather than to their left (96% of choices in the test conditions, 23/24) (see also Chapter 4). The informative condition was therefore a conceptual replication of the 1v1 condition in Legg's experiment (Legg, 2014b), the only difference being the manipulation of the females' choices through side biases instead of quinine. This change was made due to pilot trials in which the females would not take further

worms following exposure to quinine-coated worms. This behaviour may have been a consequence of their experiences with quinine in previous experiments.

In the subsequent sharing phase, males were presented with 20 choices of a single mealworm versus a single waxworm. For four of the males the choices were held by hand against the mesh of the compartment (Hoy, Romero, Caracas and Lima), for the birds who do not take food by hand (Lisbon and Dublin) the worms were placed on the wooden platform inside of the compartment. The order of presentation was pseudo-randomised such that the same type of worm did not appear on the same side on more than two consecutive choice presentations. Once the male chose a worm, the other worm was removed; if no choice was made within 30 seconds both worms were removed. Each choice presentation was followed by a 40 second inter-choice interval, in which the male could either eat, cache, or share the food with the female through the mesh window. This interval was maintained even when the male made no choice of worm. Males were required to take a minimum of 10 worms from the experimenter (50%) and were required to share at least three worms with the female across the duration of the sharing phase. If these criteria were not met in a trial, the trial was repeated on a subsequent day. The male's sharing was live-coded by the experimenter, who recorded the number of sharing events and whether the shared item was a mealworm or a waxworm. The sharing phase for each condition was also recorded with a video camera on a tripod outside of the testing compartments, and the sharing events re-checked after the fact by two coders blind to condition.

Differences in the proportion of waxworms shared[17] were assessed using two-tailed Wilcoxon signed rank tests. Comparisons were made between the two informative trials (M vs W, W vs M) and between the two uninformative trials (M vs M, W vs W) to determine whether the type of worm chosen by the female influenced males' food sharing. A further comparison was then made between the informative and uninformative conditions to determine whether informativeness of the choice affected the food sharing pattern. If the males are sensitive to the informativeness of the female's choice they should show a greater preference for sharing the worm chosen by the female in the informative condition than the uninformative condition. In contrast, if the males are responding only to the female taking a specific worm there would be no difference in the males' sharing pattern between the two conditions.

[17] The proportion of waxworms shared was derived using the formula $\frac{\text{Waxworms}}{\text{Waxworms+Mealworms}}$. The proportion of mealworms shared is the inverse of this value and therefore does not alter the output of the analysis.

6.3.3 Results & Discussion

In the informative condition, the males shared a greater proportion of waxworms when the female had chosen a waxworm (W vs M) than when the female had chosen a mealworm (M vs W) (two-tailed Wilcoxon signed rank test, n = 6, V = 21, p = 0.0313, see Figure 6.1). In contrast, there was no difference in the males' sharing pattern between the two uninformative trials (W vs W and M vs M) (two-tailed Wilcoxon, n = 6, V = 10, p = 0.625, see Figure 6.1). Furthermore, the food sharing pattern differed between the informative and uninformative conditions (two-tailed Wilcoxon, n = 6, V = 21, p = 0.0313). These findings indicate that the males are sensitive to the informativeness of their partner's choices.

This difference in sharing between the two conditions indicates that the males' sharing cannot be explained only by a simple association between the female and the eaten worm type. The female's actions towards the chosen worm were identical in both conditions (informative and uninformative), with the identity of the second worm being the only difference. If the males' sharing was based only on association between the female and the chosen worm the same pattern would be expected regardless of whether choices were informative. Furthermore, this pattern of results also cannot be explained by the female's behaviour during the sharing phase indicating her preference as the females' choices were artificially manipulated. The female's choices resulted from her side bias rather than her true preferences, meaning that any preference-indicating behaviour during sharing would not necessarily be aligned with her choices. This result therefore shows that the males take into account the identity of the other worm present at the time of choosing. Crucially, the males were able to respond to this cue in a flexible manner, altering their food sharing pattern in line with the female's choice only when it was informative.

The data from the live coding was used in this analysis; however sharing was also reassessed from the recorded video by two coders after the fact. These recordings were intended to provide a measure of inter-observer reliability but the fixed angle of the camera made it very difficult to judge which worm type was shared in the majority of the sharing events. Out of a total of 166 shared worms both experimenters were able to record the worm type shared for only 66 of them (39.8%). From the events that both experimenters coded they were in agreement on 55 of them (87.3%), however overall this means that only 33.1% of sharing events were agreed by both video coders. Furthermore, of the 55 events for which the video coders were in agreement, 7 of these agreements contrasted with the live coding.

Similar problems with video coding have been identified in previous food sharing studies with video recordings providing a clear view of just 51% of sharing events in the original study (Ostojić

et al., 2013). In this study, the authors stated that inter-observer reliability between the two video coders was high for those events that could be clearly seen, but did not actually compare the video coding to the live coding. Considering that the live coding was the source of the data used in the analysis and the results, this finding therefore does not provide any measure of the reliability of the analysed data.

However, a later study in which the live coder was blind to condition found similar results to the original study, suggesting that the live coder's accuracy is not biased by their knowledge of the experiment's predictions (Ostojić, Legg, et al., 2016; Ostojić et al., 2013). Given the issues with the video coding, the live coding appears to be the best possible option to use in the analysis. It is likely that the difficulty in reaching an agreement between the video and live coding is due to the low quality of the videos used for video coding. Thus, to address this problem, for food sharing studies conducted in the lab after this experiment (May 2018), a GoPro camera was placed inside the male's compartment to improve the viewing angle rather than using a video camera on a tripod outside of the compartment (including the Motivation Pilot in Chapter 4).

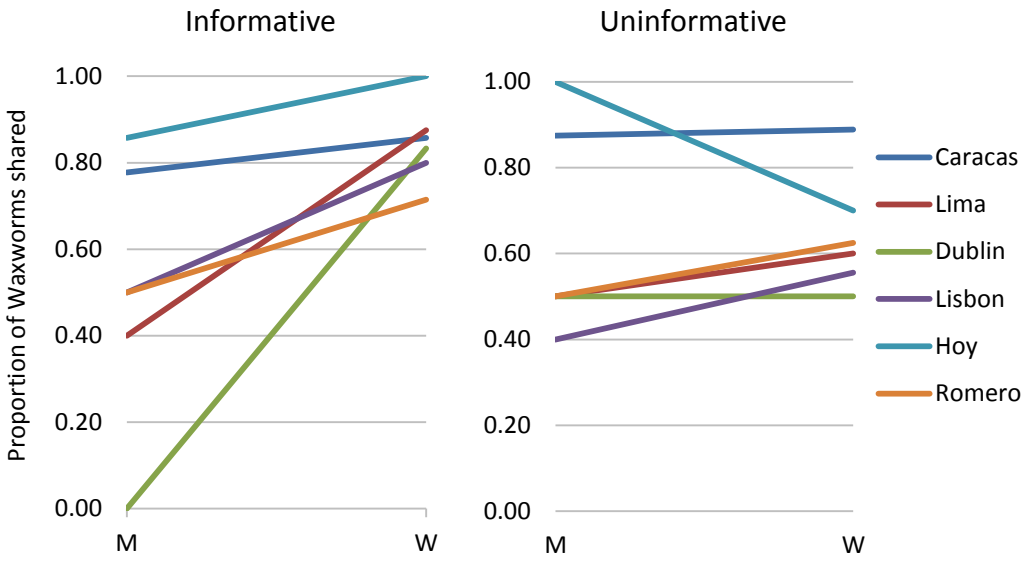

Figure 6.1 The proportion of waxworms shared for each individual bird in each test trial in the informative (left panel) and the uninformative (right panel) conditions. Paired data from each male for the trials in which the female chose mealworm (M) and in which the female chose a waxworm (W) are connected by coloured lines to show the direction of the change in the food sharing pattern.

6.4 Investigating sensitivity to informativeness of choices in human adults

Both human infants (Luo, 2011; Luo & Baillargeon, 2005, 2007) and Eurasian jays (Experiment 1) appear to show some sensitivity to the informativeness of others' choices. They appear to attribute a preference to an individual only when an object is being chosen over another object (informative choice), and not when the object is the only one available or is chosen over another identical object (uninformative choice). In order to investigate whether sensitivity to the informativeness of choices may be a shared minimal theory of mind mechanism I also conducted a comparable experiment in adult humans.

Previous investigations of implicit understanding of others' mental states have mostly used eye movement measures (see also Chapters 2 and 3). However, there have been recent criticisms of the reliability and replicability of these measures (Heyes, 2014a; Poulin-Dubois et al., 2018). In Experiment 2, I therefore chose to use a reaction time paradigm to explore participants' expectations about an agent's preferences. Participants were shown videos of an animated agent making choices between different objects by moving towards them. Participants were asked to indicate which direction the agent was travelling in. If participants extracted the agent's preference from having seen the agent choices, it is predicted that they should be slower to respond and make more mistakes when the agent subsequently violates this preference.

Reaction times have been used previously to explore adults' implicit processing of others' beliefs (Kovács et al., 2010). Some criticisms were made regarding this study, as a difference in the timing of the attention check between the conditions may have been a confound in the experimental design (J. Phillips et al., 2015). However, this should not be a problem in the paradigm I have designed here as the familiarisation videos and the timing of the response cue in the test video (the agent moving) are identical between the incongruent and congruent trials.

6.5 Experiment 2a - Pilot

In Experiment 2a I ran a pilot of the reaction time paradigm using only informative choices to identify any potential problems with the design. Participants were shown an animated agent making a series of choices between two circles of different colours (red and blue). A colour should be a simple preference rule that participants can identify easily (see Chapter 3). Participants were asked to indicate the direction the agent was travelling as quickly as possible using the arrow keys. In the final test video, the agent's choice differed between trials; in the congruent trials the agent's colour choice matched their previous preference, whereas in the incongruent trials the agent chose the other colour. If the participants implicitly attributed a

preference to the agent it was predicted that they would be slower to respond and make more errors in the incongruent trials than the congruent trials.

One specific issue I wished to test in the pilot study was whether a fully within-subject design was appropriate. Reaction times are likely to vary highly between participants, and as such a within subject design is preferable for the comparison between congruent and incongruent trials. However, to assess participants' sensitivity to the informativeness of choices a further comparison is also necessary between informative and uninformative choices. There may be issues with a fully within-subject design as there may be carry over effects between conditions. Once the participant has experienced an incongruent trial in one condition the effect in the second condition may be reduced as they may no longer be as surprised by the incongruent choice. In order to test this, I presented participants with the informative condition twice. If they showed a slower reaction time in the incongruent trial for both the first and second congruent-incongruent pairs, this would indicate that a fully within-subject design was appropriate. However, if the participants only showed a difference in reaction time in the first incongruent-congruent pair this would validate the use of the paradigm but the within-subject design would need to be changed to a mixed factorial design.

6.5.1 Participants

The final analysed sample included 19 participants, 2 further participants were tested but excluded from the analysis as they did not respond to one of the familiarisation videos. Out of these 19 participants, 12 were female and their mean age was 27 years 11 months, with an age range of 19 years 10 months to 57 years 1 month. Participants were recruited using the online Cambridge Psychology Research Sign-Up System, and paid £3 for taking part in the experiment. Participants were tested in the Department of Psychology at the University of Cambridge in June and July 2018. The experimental paradigm was approved by the Department of Psychology Ethics Committee (PEC) at the University of Cambridge.

6.5.2 Methods & Analysis

Participants were seated in front of a 27-inch screen allowing unconstrained viewing at a distance of approximately 1m. During this study videos were presented to the participants using PsychoPy2 software. Participants were first shown an instruction screen: "You will be watching a series of short videos in which a smiley face makes choices between two different objects. Your job is to react as quickly as possible to their choices. Press the left or right arrow key as soon as you know which direction they are moving. If you are ready to start press the space key."

The videos featured an animated agent (smiley face) making a choice between two circles of different colours (red and blue) by moving to one of them (Figure 6.2). Each participant was presented with four rounds of videos. Each round consisted of four familiarisation videos in which the agent consistently chose the same colour of objects. The familiarisation videos were counterbalanced so that two videos featured the preferred colour on the left side, and two on the right side. The order of presentation of the familiarisation video was randomised. The fifth video of each round was a test video in which the agent chose either the same colour, in line with their previous preference (congruent trial) or the other colour, violating their previous preference (incongruent trial). The four rounds (incongruent red, incongruent blue, congruent red and congruent blue) were presented in a random order. Each round featured an agent of a different colour to indicate that their preferences were independent across rounds.

For each video both the accuracy (correct or incorrect) and the reaction times of the participants' key presses were recorded. After the presentation of each video, including the test video, participants were given feedback on whether their answer was right, and if they had responded correctly their reaction time was also displayed. Each trial lasted 4 seconds and if no key press was made the message "Failed to respond" was displayed before moving on to the next video. The results were outputted from PsychoPy as a .csv file for each participant and collated after all participants had been tested. The reaction times were analysed using a paired one-tailed Wilcoxon signed rank test to compare between the congruent and incongruent trials, summed across the two trials of each type.

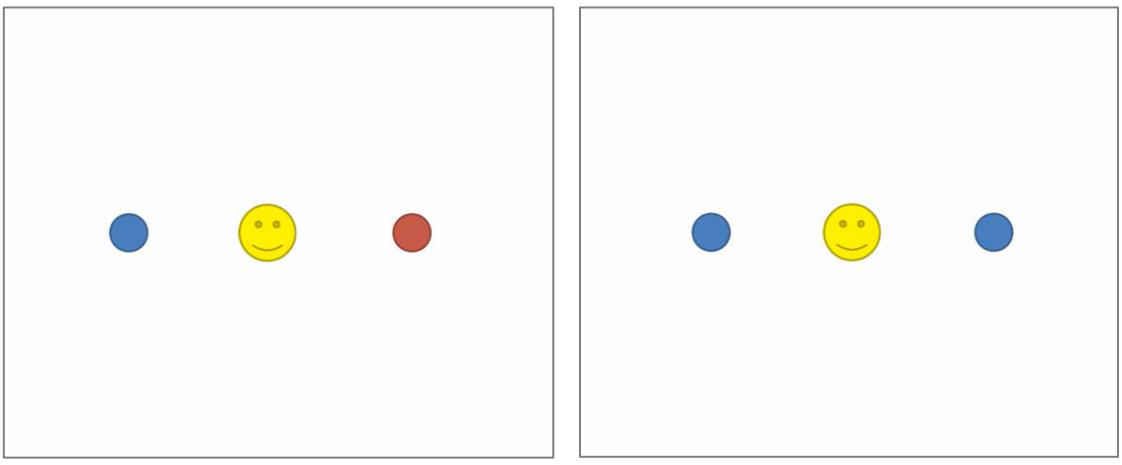

Figure 6.2 Example stills from an informative trial (left) and an uninformative trial (right). Informative trials always featured one red and one blue circle, uninformative trials featured either two red or two blue circles. The left/right positions of the two circles and the colour of the agent were randomised between trials.

If the participants were able to extract the agent's preference during the familiarisation they should expect the agent to act in line with this preference in the final test video. It was therefore predicted that they would react more slowly and make more errors in the incongruent trials than in the congruent trials. However, it is quite likely that an order effect may exist in this paradigm. After the participant's expectation is violated once they may show a reduced response in subsequent trials. I therefore also analysed the reaction times of the first pair of incongruent and congruent trials and the second pair of incongruent and congruent trials separately using a one-tailed Wilcoxon. Finally, I compared the first pair with the second pair using a two-tailed Wilcoxon to investigate whether the pattern of results differed between the two pairs. If the effect did not persist across both pairs the planned experiment would have to be changed to a mixed factorial design.

6.5.3 Results & Discussion

The sample size was too small to accurately analyse the differences in the number of errors, but observationally participants appeared to be more likely to make mistakes in the incongruent trials than the congruent trials. No participants responded incorrectly in either the first or second congruent trials, three participants (15.8%) responded incorrectly in the first incongruent trial and one (5%) responded incorrectly in the second incongruent trial.

When summed across both pairs of trials, the participants' reaction times were not slower in the incongruent trials than the congruent trials (one-tailed Wilcoxon, n = 19, V = 103, p-value = 0.384). However, when comparing only the first trial of each type, participants did react more slowly in the incongruent trial than the congruent trial (one-tailed Wilcoxon, n = 19, V = 150, p-value = 0.0129). This was not the case for the second pair of trials (one-tailed Wilcoxon, V = 58, p-value = 0.933). A further comparison showed that there was a difference between the first and second pairs (two-tailed Wilcoxon, n = 19, V = 158, p-value= 0.00945).

Together these results suggest that the piloted paradigm may be useful for identifying participants' implicit responses to the violation of an agent's preference. However, while there was an effect in the predicted direction on the first pair of trials, there was no such response in the second pair of trials. Such a carry-over effect may be due to a lack of surprise in the second incongruent trial as the participants' expectation has already been violated in the first incongruent trial. This indicates that a fully within subject design may not be appropriate for a comparison between informative and uninformative choices. If there is a carry-over effect between trials, presenting participants with both informative and uninformative conditions might lead to the masking of any effect in the condition presented second.

Thus, in order to investigate adults' sensitivity to the informativeness of choices I chose to use a mixed factorial design in Experiment 2b. The congruent and incongruent trials need to be presented in a within subject design as there is likely to be high inter-individual variation in reaction times. However, the uninformative and informative conditions need to be given to different groups of participants in a between subject design in order to prevent potential carry over effects. Using a mixed factorial design increases the required sample size making it harder to complete in person in a reasonable timeframe. Thus, the study was moved to an online format to reach a larger pool of potential subjects.

6.6 Experiment 2b

In Experiment 2b, I applied a modified version of the piloted paradigm to assess whether adults automatically process the informativeness of an agent's choices. The videos were presented to the participants online using Gorilla, an online experiment hosting system. Participants were shown an animated agent making a series of choices between either two circles of the same colour (uninformative group), or two circles of different colours (informative group). Participants were asked to indicate the direction the agent was travelling using the arrow keys as soon as they knew the answer. The final choice of the agent for both groups was always between two different coloured circles. In the congruent trial the agent's colour choice matched their previous preference, whereas in the incongruent trial the agent chose the other colour. If the participants are able to implicitly attribute a preference to the agent they should respond more slowly and make more errors when the agent's preference was violated (incongruent trials). If the participants are sensitive to the informativeness of choices this effect should be greater in the informative group than in the uninformative group.

6.6.1 Participants

Participants were recruited anonymously through Prolific Academic, a website for recruiting study participants online, and paid £1 for taking part in the experiment. A sample size of 150 was chosen to be much larger than the pilot as the paradigm was changed to a mixed factorial design. Of these 150 participants, 53 were excluded from the analysis; 3 due to video presentation errors recorded by Gorilla, 40 because they failed to respond to one or more of the familiarisation videos and 10 because they did not respond to one or more of the test videos. This gave a total of 49 participants in the uninformative group and 48 in the informative group. The experimental paradigm was approved by the Department of Psychology Ethics Committee (PEC) at the University of Cambridge.

6.6.2 Methods & Analysis

Participants were presented with videos using the cloud software platform Gorilla (www.gorilla.sc). Participants were first shown an instruction screen: “You will be watching a series of short videos in which a smiley face makes choices between two different objects. Your job is to react as quickly as possible to their choices. Press the left or right arrow key as soon as you know which direction they are moving.” This was then followed by videos similar to those used in Experiment 2a, showing an animated agent (smiley face) making a choice between two circles. In contrast to Experiment 2a participants were presented with just two rounds, instead of four. Each round of videos consisted of four familiarisation videos followed by a test video that featured either a congruent or incongruent choice.

Participants were randomly assigned to either the informative or the uninformative group. In the informative group the familiarisation videos featured choices between two circles of different colours (red and blue). In the uninformative group the familiarisation videos featured choices between circles of the same colour (both red or both blue) (see Figure 6.2). The familiarisation videos were counterbalanced so that two videos featured the preferred colour on the left side, and two on the right side. The order of presentation of the familiarisation videos was randomised. Each round featured an agent of a different colour to indicate that their preferences were independent across rounds.

For each video both the accuracy and the reaction times of the participants’ responses were recorded. After the presentation of each video, including the test video, participants were given feedback whether their answer was correct (green tick) or incorrect (red cross). If the participants did not respond within 4.5 seconds the trial was recorded as ‘no response’, and “Failed to respond” was displayed before moving on to the next video. The results were downloaded from Gorilla as a .csv file for each round and collated after all participants had been tested. The reaction times were analysed using a one-tailed paired Wilcoxon to compare the congruent and incongruent trials in each of the two groups; informative and uninformative. I also compared the difference in reaction times between the two trials between the two groups using a one-tailed unpaired Mann-Whitney U test. If participants are sensitive to the informativeness of the agent’s choices they should expect the agent to have a preference only when the choices were informative. Thus, the reaction times should be slower in the incongruent trials than the congruent trials, and this effect should be greater in the informative group than the uninformative group. It was also expected that participants should make a

greater number of errors in the incongruent trials then the congruent trials and this was analysed descriptively.

6.6.3 Results & Discussion

The participants' error rate in the informative group was descriptively in line with the predictions and similar to Experiment 2a. In the informative group two participants (4.2%) responded incorrectly in the congruent trial and six participants (12.5%) responded incorrectly in the incongruent trial. However, a similar pattern was found in the uninformative group, three participants (6.1%) responded incorrectly in the congruent trial and seven participants (14.3%) responded incorrectly in the incongruent trial. This therefore does not provide any evidence that the participants were sensitive to the informativeness of the agent's choices. Furthermore, the errors could not be analysed statistically because of a floor effect. The number of errors is very low overall and thus the variance is too low for an accurate logistic binomial regression.

The participants' reactions times were not slower in the incongruent trial than the congruent trial in the informative group (one-tailed Wilcoxon, n =48, V = 472, p-value = 0.883, see Figure 6.3). However, participants' reaction times were slower in the incongruent trial than the congruent trial in the uninformative group (one-tailed Wilcoxon, n = 49, V = 791, p-value = 0.0383, see Figure 6.3). The difference in participants' reaction times between the two trials was not greater in the informative than the uninformative group (one-tailed Mann-Whitney U, W = 879, p-value = 0.984).

The reaction times here did not provide any evidence for an implicit sensitivity to the agent's preferences as, in contrast to the pilot (Experiment 2a), participants in the informative group did not react more slowly in the incongruent trial than the congruent trial. Furthermore, in an additional contrast to the Eurasian jay findings (Experiment 1) participants appeared to be sensitive to the congruence of the agent's uninformative choices. This is a surprising finding as it does not align with a mentalistic prediction, where the participants would be expected to attribute a preference to the agent when the choices were informative but not when they were uninformative. This therefore suggests that the slower reaction times may have been caused by an alternative mechanism, rather than a violation of the participants' expectation about the agent's choices.

It is possible that the difference in the results for the informative choices between Experiment 2a and Experiment 2b could be a consequence of conducting the latter experiment online. The online study had an unexpectedly high exclusion rate, 53/150 participants (35.35) compared to the in-person pilot study 2/21 (9.5%). This was largely due to the pre-determined exclusion

criterion of excluding participants who failed to respond in one or more of the familiarisation videos. This was intended to act as an attention check to ensure that the participants had watched all the videos and was the same criterion as used in the in-person experiment. The high attrition rate of participants may be due to participants being in more distracting surroundings when taking part in the online experiment than the in-person experiment.

Furthermore, in an online setting, participants often have access to large numbers of studies and can make more money if they run through experiments quickly, so they may be less motivated to pay close attention. However, as a dedicated research tool Prolific may have fewer issues than using a non-specialist service such as Amazon's Mechanical Turk (Palan & Schitter, 2018). Furthermore, I also used Prolific in Chapter 3 and here I found that participants did appear to be attributing a simple red vs blue preference to an agent. Thus, it is not clear whether online testing was a contributing factor to the absence of the predicted response in Experiment 2b.

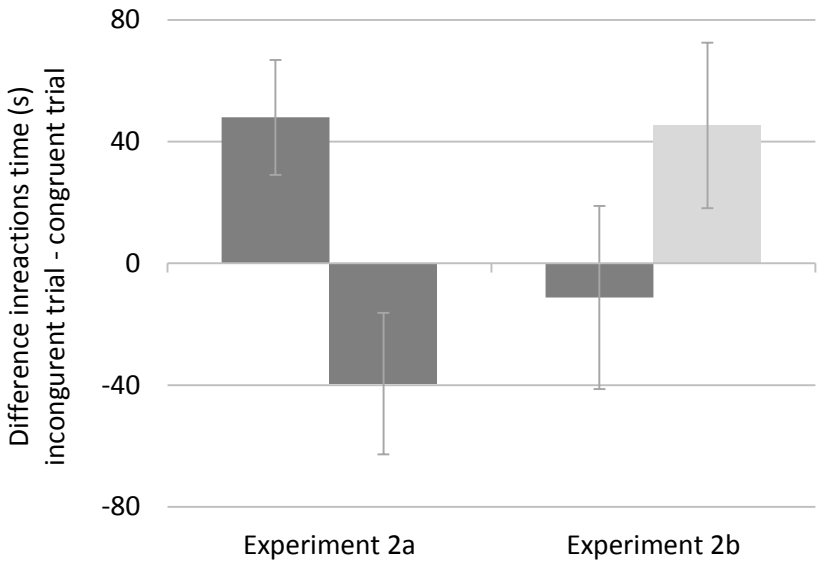

Figure 6.3 Graph showing the mean difference in reaction times between the incongruent trials minus the congruent trials across all participants. Informative trials are shown in dark grey and uninformative trials in light grey. For Experiment 2a the left bar shows the difference for the first pair, and the right bar for the second pair.

6.7 General Discussion

In this chapter I explored the sensitivity of both Eurasian jays and human adults to the informativeness of others' choices. Experiment 1 used jays' food sharing behaviour to determine whether the males attend to the choices made by their female partner. The informative condition involved a choice between two different worm types, providing information about the

female's preference for the chosen worm, whereas the uninformative condition involved a choice between two identical worms. In the informative condition males shared a greater proportion of waxworms when their partner had chosen a waxworm than when she had chosen a mealworm, but they did not do so in the uninformative condition. Thus, the results of this experiment extend the existing evidence regarding jays' food sharing responses, demonstrating that male jays attend to the choices made by their partner and are sensitive to the informativeness of these choices.

Interestingly, this response also contrasts with the response seen in previous food sharing paradigms. All previous experiments involved manipulation of the female's desires via specific satiety, and the males responded by sharing relatively *less* of the food he had seen the female eating. This paradigm involved the exact opposite response; with males sharing relatively *more* of the food he had seen the female eating. Thus, males do not respond by rigidly sharing less of the food their partner has eaten, but rather they respond differently in different scenarios. As such these results provide clear evidence for flexibility in the food sharing behaviour of male Eurasian jays.

Despite this conclusion, these results are not able to clarify which mechanism might underlie this flexible behaviour. It is possible that Eurasian jays might be able to attribute desires to their partner; nevertheless, it is also entirely possible that the male jays are responding to observable cues. Specifically, while the actions of the female in taking the worm are identical between the uninformative and informative conditions, the identity of the non-chosen worm still represents a physical difference between the two conditions. The male jay could be responding to this observable environmental cue using past experience of seeing their partner make choices between food types (a learnt rule). Or alternatively they could be responding based on a fixed innate rule that has been strongly selected due to food sharing's role in courtship (Goodwin, 1951, 1956; Lack, 1940). Neither case requires the male jay to represent their partner's desires or preferences.

Finally, the findings from Experiment 1 open up some new questions for investigating Eurasian jays' sensitivity to others' desires. Although in Chapter 5 I did not replicate the findings regarding Eurasian jays' sensitivity to *specific satiety* in the context of caching, it may be possible that cachers are sensitive to the *preferences* of observers when choosing what to cache. Conducting a test of sensitivity to informative choices in the competitive caching context may therefore represent a possible future direction for this research.

However, as discussed in Chapters 4 and 5, it remains possible that the food sharing paradigm may be a uniquely useful scenario for assessing the social cognitive abilities of Eurasian jays.

Either due to a high incentive to perform these behaviours, or a specific ability to respond to their partner while food sharing. An important question that remains to be addressed is whether Eurasian jays may have an adaptive specialisation related to courtship that contributes to their social cognitive abilities in the context of food sharing.

To complement these results in Eurasian jays, in Experiment 2 I investigated the possibility that this sensitivity to the informativeness of choices may be shared with humans. Given the evidence from infant violation of expectation studies (Luo, 2011; Luo & Baillargeon, 2005, 2007) and the results with jays in Experiment 1, it is possible that adult humans may also possess some implicit response to the informativeness of others' choices. In order to assess this, I designed a reaction time task influenced by the paradigm used with jays in Experiment 1. Participants watched videos of an agent making choices. In the informative condition an agent chose between a red circle and a blue circle, while in the uninformative condition the agent chose between two circles of the same colour. The final video was either congruent, where the agent chose the same colour as in the familiarisation videos, or incongruent, where the agent chose the other colour.

In Experiment 2a the paradigm was validated with informative choices, finding that participants were slower to respond when the agent made incongruent choices than when the agent made congruent choices. However, this effect was not present in repeated trials, suggesting that there may have been carry over effects. Thus, Experiment 2b was carried out with a mixed factorial design, testing informative and uninformative choices in a between subject format, and was also moved online in order to maximise sample size.

Contrary to the findings in Experiment 2a, participants in the informative group did not respond more slowly in the incongruent trial than the congruent trial. However, unexpectedly, the uninformative group did show a difference in reaction times between the incongruent and congruent trials. This contrasted with the prediction that the participants would attribute a preference to the agent when the choices were informative but not when they were uninformative. Observationally, the pattern of error rates in Experiment 2b also did not support a sensitivity to the informativeness of others' choices as they did not appear to differ between the informative and uninformative groups. These findings may therefore suggest that a mechanism other than implicit mentalising was driving the difference in the participants' reaction times.

One possible explanation for these findings is that they were a consequence of domain-general processes (Heyes, 2014a, 2014b). Heyes has argued that studies on theory of mind in infants that use violation of expectation or anticipatory looking paradigms could be explained simply by

differences in the degree of novelty between the conditions. Specifically, in addition to potential differences in the mental state of the protagonist, conditions also differ in low-level properties such as colour, shape or the movements of stimuli (Heyes, 2014a). Infants respond to this higher novelty by looking for longer at the stimuli, thus producing the patterns seen in existing infant studies. In a similar vein, Heyes has also criticised studies assessing implicit mentalising in adults (Heyes, 2014b). Heyes argues that it is possible that domain-general processes may simulate the effects of mentalising, but do not truly qualify as mentalising as they don't directly relate to mental states. Instead Heyes labels these processes as 'submentalising', and argues that existing experiments fail to adequately distinguish implicit mentalising from alternative submentalising explanations.

These arguments may be relevant to the findings from Experiment 2b as the low-level properties of the stimuli (e.g. colour) differ between the informative and uninformative groups. For the uninformative group, the test video is the first time in which a circle of the second colour is introduced because the familiarisation videos involve the agent choosing between two circles of the same colour (see Figure 6.2). In contrast, in the informative group both the familiarisation videos and the test videos include circles of both colours. Movement towards the other colour may have been more surprising for the uninformative group than the informative group, as the newly introduced colour had higher novelty. Thus, it is possible that the slower reaction times in the incongruent trial for the uninformative group may have been caused by domain-general processes. This experiment therefore does not provide evidence in support of an implicit sensitivity to the informativeness of others' choices in adult humans.

These findings in adult humans are doubly surprising given that both human infants (Luo, 2011; Luo & Baillargeon, 2007; Scott & Baillargeon, 2014) and Eurasian jays (see Experiment 1) have appeared to show sensitivity to aspects of others' choices. However, it is possible that the reported findings in human infants may be less robust than previously asserted. Many of these infant experiments have employed violation of expectation measures, which have been criticised repeatedly. Suggestions have been made that the findings do not truly support the authors' conclusions (Heyes, 2014a; Perner & Ruffman, 2005). In addition, as I mentioned above (Footnote 16) there has been some conflicting evidence that suggests that infants will continue to expect an agent to choose object A even when extra objects become available (e.g. A and B) (Hernik & Southgate, 2012). Proponents of the theory that infants are able to attribute preferences suggest that this is due to differences in the paradigms such as increasing the effort required to reach A (Biro et al., 2011; Scott & Baillargeon, 2014). But it is unclear whether these types of 'goal-directed' actions merely make the selection of A more salient to the infant,

therefore causing them to expect the selection of A without requiring the attribution of a preference for A over B. Overall, the evidence for implicit attribution of preferences to others' in infants may not be as strong as suggested by violation of expectation researchers, making this finding in adults less surprising.

Similarly, there is currently a 'replication crisis' facing the use of looking measures in theory of mind research with infants (see Chapter 2) (Burnside et al., 2017; Dörrenberg et al., 2018; Kulke et al., 2017; Kulke, von Duhn, et al., 2018). As a result, it is hard to determine which elements of the literature using violation of expectation or anticipatory looking measures involve reliable effects and which findings might be misleading. The increasing number of replication failures and potential criticisms of the experimental paradigms (Heyes, 2014a) may weaken the empirical evidence for implicit mentalising.

However, it must be emphasised that while Experiment 2b does not provide evidence in favour of an implicit sensitivity to choices, it also cannot provide disconfirming evidence regarding this possibility. It is always difficult to interpret negative or unexpected findings as there can be many reasons why a novel paradigm might not align with predictions even when the substantive theory is true. For example, flaws in the experimental method or underlying assumptions. It remains possible that the experiment simply failed to assess this sensitivity, as reaction times have not been used to gauge implicit mentalising in this exact manner before. In the future, this question could be explored further by repeating the experiment in person to avoid issues of distraction and high attrition. Alternatively, if reaction times are not an appropriate measure of implicit understanding in this context it might be possible to employ different measures, such as intentional action (see Chapter 3). Regardless of the results of the human experiment, the findings in Eurasian jays are an interesting addition to the body of evidence surrounding the food sharing response, providing additional indications that the males' sharing behaviour may be governed by a flexible cognitive strategy.

7. Chapter 7 – Discussion

7.1 Overview of thesis outcomes

In **Chapter 1**, I introduced some of the difficulties in researching theory of mind in non-human animals. Although it is widely accepted that human adults use mental state attribution in everyday life, there is still debate over whether non-human animals have theory of mind, and if so what form this theory of mind might take. This debate is exacerbated by disagreements over what constitutes evidence for theory of mind in animals, with sceptics promoting alternative behaviour-reading explanations for the major paradigms. In this thesis I have focused on exploring the attribution of desires and preferences, using Apperly and Butterfill's minimal system to frame my experimental approach (Apperly, 2010; Apperly & Butterfill, 2009; Butterfill & Apperly, 2013).

Despite a broadening scope in mind-reading research - encompassing great apes (Call & Tomasello, 2008), corvids (Dally et al., 2010) and dogs (Horowitz, 2011), among others - the majority of research has focussed on epistemic mental states such as belief as the strongest test for theory of mind. This mirrors the infant literature, in which various false belief tasks have been traditionally held as the gold standard for measuring theory of mind (Baron-Cohen et al., 1985; Wimmer & Perner, 1983). As a consequence, non-epistemic mental states have often been overlooked. However, given that an explicit understanding of desires emerges earlier in infant development than an explicit understanding of beliefs, exploring non-human animals' attribution of desires and preferences may have the potential to provide answers about the development and evolution of theory of mind in humans.

In this thesis I have focused on Eurasian jays as a comparator for adult humans due to the pre-existing evidence that male Eurasian jays may be sensitive to the desires of their partners while performing food sharing courtship behaviours (Ostojić, Legg, et al., 2016; Ostojić, Legg, Shaw, et al., 2014; Ostojić et al., 2013). Building on these previous studies my approach was guided by three main questions, reflecting the possibility of a shared minimal system for desire attribution in adult humans and Eurasian jays:

1. Do adult humans have an implicit sensitivity to the desires and preferences of others?
2. Do Eurasian jays have a generalised sensitivity to others' desires and preferences?
3. Are non-human animals' responses to others' desires and preferences comparable to adult humans' implicit responses?

In the first part of my thesis I focussed primarily on Question 1, exploring the possibility of an implicit response in human adults to the desires of others. In **Chapter 2,** stemming directly from the existing work on food sharing in Eurasian jays (Ostojić, Legg, et al., 2016; Ostojić, Legg, Shaw, et al., 2014; Ostojić et al., 2013), I investigated whether human adults may have an implicit sensitivity to others' specific satiety. I found that human adults did not consistently produce directed anticipatory looking responses when watching a sated individual make choices between foods. In Experiment 1, the human participants looked randomly at the sated and non-sated food. However, in Experiment 2, which included a verbalisation of the protagonist's satiety after eating, participants were more likely to look first at the non-sated food. It is possible that the lack of an anticipatory looking response in Experiment 1 was due to the participants attributing a general satiety to the protagonist, rather than specific satiety to the eaten food type. Thus, in Experiment 3, I assessed whether adult humans would show anticipatory looking in response to a protagonist sated on a single food type. Participants did not anticipate that a protagonist would choose to eat a smaller amount food when sated, providing no evidence to suggest that the lack of anticipatory looking in Experiment 1 was due to the attribution of general satiety.

In all three experiments, participants were able to make correct explicit predictions based on the protagonist's satiety, despite their variable implicit responses. The directed anticipatory looking in Experiment 2 resulted from the presence of a behavioural cue for the protagonist's desire – the verbalisation "I am full of apples/crackers". This cue would allow the participants to respond based on their previous experience of people's behaviour following such verbalisations, without necessarily inferring an unobservable mental state. It is therefore unclear whether the participant's looking response resulted from a minimal implicit system or simply from a 'behaviour-reading' response to observable cues.

In **Chapter 3**, I extended this investigation of a potential implicit response, using a novel paradigm based on transitive inference. I found that adult humans may show a non-associative implicit response to the transitivity of others' preferences. Previous studies have used violation of expectation paradigms to test infants' spontaneous transitive inference abilities (Gazes et al., 2017; Mascaro & Csibra, 2014; Mou et al., 2014). However, given the current questions about the validity of eye movement measures in the theory of mind literature (Kulke, von Duhn, et al., 2018; Poulin-Dubois et al., 2018; Powell, Hobbs, Bardis, Carey, & Saxe, 2018), I chose instead to use a different measure to assess adults' implicit transitive inference of preferences. This involved showing videos of an agent making a choice between objects and then asking the participant to indicate which door they would like to open for the agent, with one leading to a preferred object and one leading to a less preferred object. This measure was validated by

testing participants' responses when the agent's preferences followed a simple colour rule, which verified that the participants' answers aligned with the agent's preferences.

In Experiment 1, I tested participants on a preference hierarchy with 5 terms (A>B>C>D>E), and found that the participants did not show an implicit sensitivity to the agent's preferences. Thus, given the potential memory load of a 5-term hierarchy, in Experiment 2a I tested participants on a simpler 3-term hierarchy (A>B>C). Here participants' choice of door indicated that they were sensitive to the preference of the agent. Crucially, in Experiment 2b I found that this sensitivity did not appear to be driven by an association between the agent and the chosen objects.

In all experiments I also included an awareness question that asked participants to rank the objects according to the agents' preference. The majority of participants were unable to rank the objects correctly, suggesting that they did not have explicit knowledge of the full hierarchy. However, participants showed a tendency to rank the final element of the hierarchy (E or C) at a lower position relative to the other objects. The test pair in the 3-term hierarchy (A vs C) included this final element, so it is not entirely clear whether the participants' answers involved logical reasoning about the least preferred object, or were due to a form of implicit transitive inference that is both non-cognitive and non-associative.

In the second part of my thesis I engaged with Question 2, assessing whether Eurasian jays may have a sensitivity to others' desires outside of the food sharing context. In **Chapter 4**, I explored the generalisability of jays' co-operative social cognitive abilities by designing new paradigms to assess their sensitivity to their partner's specific satiety. In doing so I encountered some of the issues researchers typically face when trialling novel paradigms. In Experiment 1, I designed a paradigm building on previous experiments that have used looking measures to assess social cognition, in corvids, humans and great apes (Bird & Emery, 2008; Krupenye et al., 2016; Scheid et al., 2007; Scott & Baillargeon, 2017). However, the validation stage of this paradigm failed as the jays did not look through the peepholes with high enough frequency to realistically reveal an effect. Furthermore, when they did look through the peepholes, their choice was compromised by a side bias – an underlying preference for looking through the peephole on the left side. In Experiment 2, I developed a paradigm that included the provision of food to their partner in an effort to increase the jays' motivation. This involved an apparatus where the pulling of a string allowed food to swing into their partner's compartment. However, I found that the jays' motivation was again an issue - the females did not take food from the apparatus when sated and the males reached extinction on the necessary string-pulling behaviour.

Taken together, these studies highlight the difficulties of testing non-human animals in novel scenarios, particularly those in which they are required to express behaviours that they would

not perform in their natural environments. Similarly, in the past, researchers have argued that there may be issues with attempting to test animals in paradigms with low ecological validity (Hare, 2001; Hare & Tomasello, 2004). Such paradigms might require 'unnatural' behaviours, and the subjects may therefore have low motivation to attend to scenarios or interact with apparatus appropriately (Hare, 2001). As such, it is possible that the natural, highly incentivised food sharing behaviour may represent a uniquely useful paradigm for assessing social cognitive abilities in Eurasian jays. This may reduce the issues with floor effects and low engagement found in both Chapter 4 and Chapter 5 described below. However, there may be issues with assessing cognition in familiar contexts, as the subjects are more likely to have previous experience that can be used to directly predict others' behaviour. Furthermore, ecologically valid scenarios with strong selection pressure are the very scenarios in which animals are most likely to have evolved specific innate behavioural responses. It is therefore possible that the response of male jays is an adaptive specialisation unique to courtship food sharing.

Despite this possibility, existing studies seem to suggest that jays may also respond to others' desires in the competitive caching context (Ostojić et al., 2017). In **Chapter 5**, I attempted to extend the findings relating to Eurasian jays' sensitivity to the desires of a conspecific observer while caching. In Experiment 1, I used a modified inference paradigm based on a previously conducted unpublished food sharing study (Ostojić, Legg, Mendl, et al., 2014). The food the cacher saw entering the observer's compartment during pre-feeding was manipulated such that it either matched or conflicted with the actual pre-fed food. This should ensure that any behavioural cues produced by the observer (based on the actual pre-fed food) should have matched or conflicted with the information available to the cacher (the food seen entering). However, the cacher's caching pattern did not differ in line with the observer's satiety in either the matching or conflicting condition. This was despite the matching condition being a conceptual replication of the original findings (Ostojić et al., 2017).

I therefore attempted to directly replicate both experiments of the original study; a pilfering experiment (Experiment 2a) and a caching experiment (Experiment 2b). I found that the observers' pilfering pattern did not appear to be influenced by their own satiety, and that the cachers' caching pattern did not appear to be influenced by the observer's satiety. Thus, either my replication attempts yielded false negative results or the original findings in Eurasian jays were false positives. It is possible that the low sample size available may have contributed to a false negative in this replication. The original study may have given an overestimate of the effect size and thus the replication was underpowered, despite including at least as many Eurasian jays as the original study (Gelman & Carlin, 2014). The low level of both pilfering and caching from

the birds could have caused a floor effect making it difficult to determine an effect even if it were present; increasing the likelihood of a false negative. However, alternatively, it is possible that these floor effects combined with a small sample size may have caused a false positive result in the original study.

This study may highlight some general issues within the field of corvid research (see Section 5). Few corvid cache protection studies have been directly replicated, and where repeated studies have been published they have often been conducted in different species. In contrast, the finding that male jays alter their food sharing in line with the specific satiety of their partner has been directly replicated in a number of studies both published and unpublished (Ostojić, Legg, et al., 2016; Ostojić, Legg, Mendl, et al., 2014; Ostojić, Legg, Shaw, et al., 2014; Ostojić et al., 2013). It therefore remains possible that Eurasian jays' sensitivity to the desires of their partner may be specific to the food sharing context, or at the very least may be more consistently detectable in this context due to a higher motivation on the behalf of the birds.

As a consequence of these findings in Chapters 4 and 5, in **Chapter 6** I returned to the food sharing context. Here, I compared Eurasian jays' and humans' sensitivity to the informativeness of others' choices using a novel paradigm. In Experiment 1, I found evidence that in the food sharing context Eurasian jays respond differently depending on whether the saw their partner make an informative or uninformative choice between two food items. In the informative condition, where the female chose between a mealworm and a waxworm, male jays shared a greater proportion of waxworms when she chose the waxworm than when she chose a mealworm. No such pattern was seen in the uninformative condition when the female chose between two waxworms or two mealworms. The difference between these two conditions indicates that male jays are sensitive to the informativeness of their partners' choices in the food sharing context. This suggests that males are able to respond flexibly while food sharing, rather than following a rigid rule that is only valid in the specific satiety paradigm.

Subsequently, in Experiment 2, I attempted to find comparable evidence for an implicit sensitivity to the informativeness of choices in adult humans. I employed a reaction time measure, anticipating that participants would be slower to respond and make more errors when their expectation of an agent's preference was violated. In the pilot study, I found that participants' reaction times differed based on whether an agent's movements were congruent or incongruent with their preferences. In order to maximise the sample size I conducted an online experiment, which compared participants' reaction times and error rates between congruent or incongruent trials, when the agent's choices were either informative or uninformative.

However, in the informative condition I found no effect of congruence on the participants' reaction times. Furthermore, in contrast to the predictions of the experiment, in the uninformative condition the participants' reaction times were slower in the incongruent trial than the congruent trial. Thus, overall the experiment did not provide evidence in favour of an implicit system sensitive to the informativeness of others' choices. Instead these results may favour a low-level explanation, such as a response to differences in novelty (Heyes, 2014a).

7.2 Implicit sensitivity to others' desires in human adults

Together the results of my experiments with human adults may provide tentative support for an implicit sensitivity to the desires and preferences of others. However, it is unclear what mechanism or mechanisms might underlie their responses in each scenario. In Chapter 2, I found that adults only showed a directed anticipatory looking response when the specific satiety of the protagonist was verbalised. It is therefore possible that this spontaneous response was driven by a direct reaction to this observable cue as opposed to an unobservable mental state. This finding therefore does not appear to support the suggestion of a minimal system of theory of mind sensitive to the mental states of others. In Chapter 3 I found that adults may be able to respond to the transitivity of an agent's preferences when they lack an explicit representation of the *full* hierarchy. However, it is possible that participants' responses to the test pair question were driven by a conscious representation of the least preferred object rather than a minimal system that is sensitive to preferences or desires.

Finally, in Chapter 6, I found that human adults did not show the expected implicit sensitivity to the informativeness of others' choices in my reaction time paradigm. Contrary to the predictions of the experiment, participants appeared to be sensitive to the congruence of an agent's choices only when these choices were uninformative and not when they were informative. This finding was inconsistent with predictions based on implicit mentalising and may be more appropriately explained by domain-general processes such as low-level novelty (Heyes, 2014a, 2014b). Overall, the studies presented here in this thesis do not provide convincing evidence in favour of a minimal system that automatically and consistently tracks the desires of others.

This may be a surprising finding given that there appears to be a large amount of existing evidence in favour of implicit tracking of beliefs in both infants and human adults. But, as I discussed briefly in Chapter 2, replication issues have recently been raised in infant theory of mind research. Theories suggesting an implicit theory of mind have resulted from findings utilising the anticipatory looking and violation of expectation paradigms. However, there have been an increasing number of studies in which such paradigms have failed to replicate, even in

some cases with much larger sample sizes (Burnside et al., 2017; Dörrenberg et al., 2018; Kulke et al., 2017; Kulke, von Duhn, et al., 2018).

These repeated replication failures come at a time when questions are being asked across scientific fields about the way in which research is being conducted. Historically, journals have been reluctant to publish replications of existing research (Martin & Clarke, 2017), and perhaps more insidiously, replication efforts have often been derided, with researchers who conduct replications dismissed as unimaginative, or being accused of actively hindering scientific progress (Stevens, 2017). Furthermore, in smaller fields such as theory of mind, fewer research groups work on each topic decreasing the likelihood of independent replications being conducted. As stated by Poulin-Dubois et al., "it is particularly unfortunate that in the case of the [violation of expectation] paradigm, out of 15 publications on false belief, 11 (73%) have been conducted by Baillargeon and her former students" (pg. 303) (Poulin-Dubois et al., 2018).

Now that replications are being attempted, potential 'replication crises' have been identified across fields (Button et al., 2013; Munafó, 2009; Prinz, Schlange, & Asadullah, 2011) including psychological science where a major replication effort of 100 studies found that only 39% replicated the original results (Open Science Collaboration, 2015). In the specific case of theory of mind research, the failure to replicate looking measure experiments in infants raises questions about the empirical basis for Apperly and Butterfill's minimal theory of mind system.

Some researchers have tried to explain away failed replications, suggesting that implicit theory of mind may be difficult to show experimentally because the stimuli don't appropriately evoke a response or are not ecologically relevant (Krupenye et al., 2016; Kulke, von Duhn, et al., 2018). However, this attitude falls into the trap of determining the validity of an experimental method based on the results achieved. Often tests are only considered valid if they get positive results, and if the test gets negative results it is assumed that the test itself must have been invalid rather than the theory. Thus, this line of argument prevents disconfirming evidence from ever being considered valid, hindering any possibility of falsification.

However, the authors of replication studies have also directly raised questions about the validity of looking measures, and whether they are truly indicative of implicit theory of mind at all (Dörrenberg et al., 2018; Kulke, Reiß, Krist, & Rakoczy, 2018; Kulke, von Duhn, et al., 2018). There are two possible outcomes for the validity of looking measures:

(i) These looking measures *are not* a valid test of implicit theory of mind – experiments employing these measures are therefore unable to provide evidence for implicit theory of mind regardless of whether they produce positive or negative results.

(ii) These looking measures *are* a valid test of implicit theory of mind – experiments employing these measures could provide confirming or disconfirming evidence for theory of mind depending on whether they produce positive or negative results.

Currently there seems to be increasing evidence for the former option – looking measures are not a valid test of implicit theory of mind. Validity is broadly broken down into two categories: internal validity, the cause-effect relationship between the dependent and independent variables, and external validity, the generalisability of this result to scenarios which should show the same cause-effect relationship.

The internal validity of implicit theory of mind tasks experiments has been challenged (Heyes, 2014a), by suggestions that the positive findings from these looking measures may have been caused by factors other than implicit theory of mind. In these tests the dependent variable is the looking measure, while the independent variable is the mental state of the actor in the videos (e.g. true belief or false belief). For the experiments to be a valid test of implicit theory of mind any differences in the looking behaviour of the infants must be due to differences in the actor's mental state. However, criticisms of this assumption have been raised, with suggestions that the results found in these studies could be due to artefacts such as low-level novelty arising from visual differences between the conditions (Heyes, 2014a; Sirois & Jackson, 2007). Alternatively, other researchers have suggested that the infants' looking responses could result from the tracking of others' behaviour rather than their mental states (Perner & Ruffman, 2005). Finally, there have also been similar criticisms of the validity of adult looking studies used to provide evidence for Apperly and Butterfill's minimal system (Heyes, 2014b; J. Phillips et al., 2015). Thus, the internal validity of looking measure experiments may be flawed even in those studies that do produce positive findings.

Additionally, replication studies of implicit theory of mind tasks have assessed the *convergent validity* of the different studies they attempted to replicate. If different tasks are all assessing the same mechanism - implicit theory of mind - they should produce results that correlate across tasks. Explicit theory of mind shows strong correlation between tasks (Perner & Roessler, 2012; Rakoczy, Bergfeld, Schwarz, & Fizke, 2015). However, none of the studies replicating implicit theory of mind experiments have been able to find correlations between tasks, suggesting that they lack external validity (Dörrenberg et al., 2018; Kulke, Reiß, et al., 2018; Kulke, von Duhn, et al., 2018).

Overall, the criticism of the validity of these tests leads to difficulty in the interpretation of failed replications. As mentioned above (i), if looking measures are not valid tests of implicit theory of

mind this means that previous positive findings cannot be considered as *confirming* evidence for implicit theory of mind, but this also equally means that negative findings from failed replications cannot be considered as *disconfirming* evidence for implicit theory of mind. As such we are left with an absence of evidence, and as the common aphorism goes: 'absence of evidence is not evidence of absence'. However, theories such as Apperly and Butterfill's two systems account of theory of mind have relied heavily on looking measure experiments for their formulation and to provide confirmatory evidence for their mechanisms. As such, although the question of whether an implicit theory of mind may exist remains unanswered, the empirical basis for existing theories may not be as strong as previously believed.

7.3 Flexible sensitivity to other's desires in Eurasian jays

In my experiments with Eurasian jays, the findings suggest that jays' sensitivity to the desires and preferences of conspecifics may be more limited than previously believed. There are two possible explanations for these findings, either (i) their social cognitive abilities may be specifically limited to the food sharing context, or (ii) these abilities may be more clearly expressed in food sharing paradigms due to jays' higher motivation to attend to and interact with their partner in this scenario. In Chapter 4, I was unable to find a viable alternative paradigm in which jays might demonstrate sensitivity to their partner's specific satiety. This may have been caused by issues with ecological validity, resulting in the jays not performing the behaviours of interest. Similarly, in Chapter 5, floor effects stemming from a low motivation to cache may have been a contributing factor in the non-replication of jays' sensitivity to an observer's specific satiety in the caching context.

However, the findings presented in this thesis are also compatible with the hypothesis that male Eurasian jays' response to their partner's desires is specific to beak-to-beak food sharing. The previously identified sensitivity to an observer's satiety during caching could be a false positive result; and Eurasian jays may be unable to employ these social cognitive strategies with individuals other than their partner and/or outside the food sharing context. Although corvids have complex social relationships, researchers have argued that these differ from primates' social interactions in that they appear to be focused more on strong pair bond relationships rather than tracking a large number of individuals (Emery, Seed, et al., 2007). The authors state that "such forms of social knowledge appear to be used primarily within the context of the pair bond, rather than applied to a larger social network" (pg. 502). This 'relationship intelligence' might be a driver of the evolution of intelligence in birds with monogamous life-bonds. It is therefore possible that corvid species that form strong pair bonds but have limited relationships

with other conspecifics, such as Eurasian jays, might possess social abilities specifically adapted for, and limited to, partner interactions.

Nevertheless, questions remain about whether the food sharing response requires 'mind-reading', or can be explained based on differences in observable behavioural or contextual cues. Published food sharing studies have suggested that the females' behaviour at the time of sharing is insufficient to drive the difference in food sharing (Ostojić et al., 2013). However, it remains possible that the males may be sensitive to the female's behaviour during the pre-feeding stage. For example, if the female stops eating during pre-feeding and rejects the remaining food in the bowl, this behaviour may provide an indication of the female's likelihood of avoiding that food in the future. Thus, the male may respond to this behaviour directly, sharing less of this food with the female based on either a learned or innate rule.

However, whether or not this response falls under 'behaviour-reading', this does not diminish the fact that male Eurasian jays demonstrate a strong sensitivity to their partner in the food sharing context. In Chapter 6, I found evidence that male jays' ability to respond to their partner's desires is flexible and not just a rigid response that can only be applied to specific satiety. Males were not only able to adapt their response to share food in line with their partner's preferences, but were also sensitive to the informativeness of these choices. These responses appear to require flexible social cognition, however it remains unclear whether male jays' food sharing involves desire state attribution or simply a response to the female's observable behaviour.

This conclusion may seem surprising given the substantial amount of literature promoting the idea of theory of mind and advanced social cognition in corvids (Dally et al., 2010; Emery & Clayton, 2004; Emery, Dally, & Clayton, 2004; Seed et al., 2009). However, I argue that comparative cognition research is often biased towards attributing cognitive abilities where experimental evidence is insufficient or ambiguous. This is likely the result of a combination of the way research is conducted in comparative cognition and the inherent limitations specific to the field (Beran, 2018; Stevens, 2017).

Firstly, researchers working in animal cognition often have to contend with limited sample sizes due to a small number of available captive individuals. As such, direct replications of findings are rare, as finding a sufficiently large independent sample in which findings can be replicated is difficult, if not impossible. Additionally, as the scope of comparative cognition has extended, a more diverse range of species has been tested, and this has had negative consequences for the ability to replicate studies. Since Beach raised the issue of focussing on a small number of species in his classic paper "The Snark was a Boojum" (Beach, 1950), comparative psychology

has embraced the study of a range of species from corvids to cephalopods (Shettleworth, 2009). While this reflects positive progress in the avoidance of anthropocentrism and a greater breadth of research questions, as a consequence of this diversity some species are being assessed by a very small number of researchers, and in some cases a single research group.

A prime example of this is corvid cognition, in which research groups tend to focus on a single species and there is often little overlap in species tested between groups. For example, currently all research testing cognition in Eurasian jays is being carried out with birds from a single research group at the University of Cambridge. If there is no other research group studying the same species, it is not possible for independent replications to be conducted. In addition, corvids are long-lived and usually kept in captivity throughout their lives. The same individuals are therefore often used in repeated studies and tested for upwards of ten years. If a research group repeats their own studies, this will often involve re-using subjects from the original study in order to achieve a sufficiently large sample size (see Chapter 5).

Similarly, in corvid research, where paradigms are repeated, they often use a different species either because researchers don't have access to the original species or because testing the paradigm with a new species makes their results more 'novel' and easier to publish. As a consequence of this practice, if the researchers do not replicate the original results this can be explained away by species differences, without considering the possibility of the original result being a false positive.

Finally, even when studies can be replicated in the same species, low sample sizes mean that the replication study will likely lack the power to detect meaningful effect sizes. Thus, a replication study might fail to return a significant effect, even where the effect is true, making it difficult to interpret the results of a replication (see Chapter 5). As I discussed in Chapter 4, this issue can be compounded when the behaviours of interest occurs at low frequency with high noise, making it even harder to detect a true effect in low-powered studies. Thus, researchers may be reluctant to replicate their own findings as there is a strong possibility of failing to find the same effect, undermining their own body of research.

Further adding to this replication issue is the fact that there is a directional bias within the structure of comparative cognition. Research programmes typically focus on determining whether a study species possesses a specific cognitive ability. Thus, studies are often structured such that the null hypothesis assumes that the subject *does not have* the ability in question; while the alternative hypothesis assumes that the subject *does have* the cognitive ability in question. In null-hypothesis significance testing, positive results provide evidence in favour of the alternative hypothesis, whereas negative results can be difficult to interpret and often do

not provide direct evidence in favour of the null hypothesis (Cohen, 1994; LeBel & Peters, 2011). Furthermore, negative findings are often dismissed post-hoc by extraneous factors, such as flaws in how the experiment was conducted or designed, rather than considering the results to be a true negative finding (see discussion of validity above). As a consequence of this, research programs are often directionally biased towards finding evidence in favour of a cognitive ability rather than evidence against it. This drives research in the direction of accumulating positive evidence for advanced cognitive abilities, regardless of whether that cognitive ability is in fact present in that species.

This can then be worsened by confirmation bias on the behalf of the researchers. For example, researchers might allow seemingly contrasting results to be used to support the same conclusion (van der Vaart & Hemelrijk, 2012). In one study, scrub jays were found to cache in far locations when an observer had visual access but not when their view was blocked by a screen (Clayton, Dally, Gilbert, & Dickinson, 2005). In contrast, in another study Clark's nutcrackers were found to perform cache protection strategies in both an observed condition, where the observer could see the caching, and an in-private condition, where the observers' view was blocked (Clary & Kelly, 2011). In the former study, it was suggested that this sensitivity to visual access indicates an understanding of the observer's perspective, while in the latter, it was suggested that the birds could generalise their cache protection response based both on visual and auditory cues indicating an observer. Thus, in both cases, this conflicting evidence was used to conclude that the birds in question were employing complex cognition in order to protect their caches.

Confirmation bias can be particularly insidious in comparative cognition as there is a substantial distance between the theoretical claims and the actual statistical hypotheses tested in our experiments, thus allowing flexibility in interpretation. In comparative cognition, researchers are always attempting to test unobservable factors, specifically the unobservable cognitive processes that underlie observable behaviours. Thus, to construct experimental paradigms, the observable variables being measured must be linked to these unobservable variables by the way of auxiliary assumptions.

As Trafimow suggests, "one might measure attitudes by observing where check marks have been made on an attitude scale, under the assumption that the placement of the check marks has something to do with attitudes in people's minds that cannot be observed directly" (Trafimow, 2012). Such auxiliary assumptions must be true in order for the statistical hypothesis to be a valid test of the substantive theory. Thus, if there is substantial distance between a theory and the statistical hypothesis, when a negative result is obtained it is not clear whether the null finding arises from a failure in one of the auxiliary assumptions or because the theory itself is

false. If researchers are strongly attached to their own theories, they can appeal to flaws in any number of these auxiliary assumptions when their results do not match their prediction (Meehl, 1967).

In the case of the food sharing studies the theoretical claim might be that male jays use desire state attribution to respond to their partners' specific satiety. However, to extract this from the observable measured variable - the proportion of waxworms shared by the male - a number of assumptions must be true. Namely, the pre-feeding of the female must have led to specific satiety, the male must have attended to the female's satiety, the male must have attributed a desire to the female based on this satiety, and the male must have used this attributed desire to alter the food they chose to share. If the males do not alter their food sharing in line with the female's satiety this could be because they are unable to attribute desires, but it could also be because the male did not attend to the pre-feeding or because the females did not eat enough of the pre-feeding food to appear sated.

Furthermore, while these assumptions may be valid explanations for a positive finding, there are also alternative possibilities. For example, the female may alter which food she chooses to take from the male during sharing, or the male may respond directly to the female's behaviour at the time of pre-feeding without attributing a desire. Thus, while positive results from this experiment may be consistent with a desire state attribution theory, they are not sufficient to prejudice this theory over other equally valid explanations. The greater the number of auxiliary assumptions the more difficult it is to draw a direct line between the substantive theory and the findings, making the results more ambiguous and opening the way for the sort of methodological criticisms common in theory of mind research (Heyes, 2015; Penn & Povinelli, 2007).

Furthermore, while theory of mind researchers might mention these alternative 'behaviour-reading' explanations in their papers, they tend to focus predominately on potential mind-reading explanations for the obtained results. For example, when discussing scrub jays' cache protections strategies Dally et al. mention that "keeping track of who was watching when need not require a humanlike "theory-of-mind" in terms of attributing unobservable metal states, but... may result from behavioral predispositions in combination with specific learning algorithms" (pg. 1665) (Dally et al., 2006). Similar qualifications are present in other scrub jays papers (Dally et al., 2005), however, as highlighted by Penn and Povinelli, the authors still conclude that the best or most parsimonious explanation is that scrub jays are able to attribute the mental states of 'seeing' and 'knowing' to observing conspecifics (Penn & Povinelli, 2007). Moreover, when reviews are compiled of such studies, these alternative behavioural

explanations tend to be entirely overlooked in favour of accounts detailing how these findings are consistent with theory of mind (Clayton et al., 2007; Dally et al., 2010; Emery et al., 2004).

Adding to this issue is the fact that, as well as avoiding replications (see Section 3), journals are often biased towards publishing novel, flashy or unexpected findings. Negative findings, which are seen as difficult to interpret, remain unpublished, resulting in a 'file drawer effect' (Ferguson & Heene, 2012; Rosenthal, 1979). As Povinelli and Vonk put it, "experiments that would appear to confirm the presence of theory of mind abilities are hailed as 'breakthroughs' whereas our own, carefully conducted, programmatic set of over two dozen studies ... is dismissed as part of a general pattern of 'negative evidence'" (p. 17) (Povinelli & Vonk, 2004).

Moreover, as both the researchers' future career and their respect in the field rely on producing positive outcomes, these selection pressures can also favour research practices that increase the likelihood of false positive results. Researchers often intentionally or unintentionally engage in 'questionable research practices' (QRPs) that can inflate their likelihood of positive findings (Simmons, Nelson, & Simonsohn, 2011). Specifically, it is a common issue that aspects of the experimental method are not clearly defined prior to conducting an experiment (Ferguson & Heene, 2012). For example, failure to define sample sizes, exclusion criteria, the key behaviours of interest or the analysis strategy in advance (Simmons et al., 2011). This increased methodological flexibility can result in analytical decisions that favour significant p-values ('p-hacking') or the presentation of post-hoc conclusions as confirmatory research ('HARKing' - Hypothesising After the Results are Known) (Kerr, 1998; Wagenmakers, Wetzels, Borsboom, & Van Der Maas, 2011). Although estimates of the prevalence of QRPs in psychology vary, it is agreed that the usage of such practices is non-trivial (Fiedler & Schwarz, 2016; John, Loewenstein, & Prelec, 2012).

Overall, a combination of external incentives to produce certain types of results and internal confirmation bias stemming from a biased literature can lead researchers to interpret their results in a way that supports their chosen theory. As a consequence of this, it is difficult to distinguish whether progress in the field results from improvement in methods, or whether so-called 'breakthroughs' might be the result of biased reporting of false positive findings (Meehl, 1967). False positive findings are a necessary consequence of significance testing, however file drawer effects, directional biases and QRPs can exacerbate the issue, making false positives more common (Ioannidis, 2005). As such, the cognitive abilities of non-human animals are likely to be overestimated within the comparative cognition literature, causing researchers to develop a skewed view of their study species' abilities (Nissen, Magidson, Gross, & Bergstrom, 2016). This further promotes the development of a biased literature, which appears to present strong

evidence in favour of the prevailing theory (in this case corvid theory of mind) while the studies themselves may be un-replicated and frequently over-interpreted. Hence, I suggest that my findings in this thesis may be less surprising than they appear at first glance; reflecting more closely the reality of corvids' social cognitive abilities.

7.4 Shared mechanism for responding to others' desires

Finally, the driving question behind my thesis was whether there may be a shared minimal form of theory of mind that would allow both corvids and humans to track and predict the behaviour of others. Proponents of corvid intelligence have argued that there is extensive convergence of complex cognition in corvids and great apes (Emery & Clayton, 2004; Seed et al., 2009; Van Horik et al., 2012). Corvids and apes appear to share a number of traits, such as a large relative brain size, cognitive and behavioural flexibility, and complex social relationships. Despite the differences between the social structures of corvids and primates (Emery, Seed, et al., 2007), sociality may still be an important factor for intelligence in both groups, and social cognitive abilities have been frequently cited as a potential area of convergence. Specifically, it has been suggested that corvids, like apes, demonstrate social cognitive abilities that may involve sensitivity to the perspectives, knowledge and desires of others (Clayton et al., 2007; Ostojić et al., 2013).

In proposing their minimal system for theory of mind, Apperly and Butterfill state that this model may represent a fundamental mechanism shared by humans and non-human animals. Specifically, that "the success of infants and nonhuman animals on some belief reasoning tasks may be best explained by a cognitively efficient but inflexible capacity for tracking belief-like states" (pg. 953) (Apperly & Butterfill, 2009). I focused on desires as a potential area for a shared sensitivity to others' mental states for two main reasons. Firstly, because an understanding of desires may be representationally simpler and earlier developing than an understanding of epistemic states, such as beliefs, making it more likely that this ability may be present in non-human animals. And secondly, because male Eurasian jays had already been found to show some sensitivity to the desires of their partner in the food sharing context (Ostojić, Legg, et al., 2016; Ostojić, Legg, Shaw, et al., 2014; Ostojić et al., 2013).

In Chapter 2, using a paradigm influenced by the Eurasian jay studies, I assessed whether adult humans may have a similar sensitivity to others' specific satiety, at the implicit level. I found that participants automatically anticipated the actions of a sated protagonist, but only when the protagonist directly verbalised their satiety. This result may be somewhat surprising given the evidence for Eurasian jays' sensitivity to the satiety of others (Ostojić et al., 2013). However, as

described above, male Eurasian jays in this paradigm may be responding to behavioural cues produced by their partner during pre-feeding, rather than responding based on her mental states. Thus, it is possible that the anticipatory looking response in humans and the sensitivity to satiety in jays may both be driven by behavioural cues. However, despite this possible similarity, if human adults do not show a consistent implicit response to others' satiety then this does not appear to provide evidence for a generalised minimal theory of mind that automatically processes others' desires.

Similarly, in Chapter 3 I investigated adult humans' implicit sensitivity to the transitivity of others' preference. Transitive inference is an ability shared by humans and some non-human animals, and thus it may be a candidate for convergent cognition. Transitive inference is thought to be related to sociality and is usually tested in animals in the context of tracking dominance relationships between members of a group. There is pre-existing evidence for a spontaneous response to the transitivity of preferences in infants (Mou et al., 2014) and I found some evidence for sensitivity to the transitivity of others' preferences in human adults. However, this does not yet provide any evidence for a shared minimal system, although it might be worth further investigation, including exploring the possibility of a sensitivity to the transitivity of preferences in Eurasian jays.

Finally, in Chapter 6 I conducted a comparative study, assessing whether Eurasian jays and human adults are sensitive to the informativeness of others' preferences. Male Eurasian jays appeared to show a flexible ability to respond to the choices of their partner when food sharing, including a sensitivity to the informativeness of these choices. However, in the human experiment I found an unexpected difference in participants' reactions times in response to uninformative choices, suggesting that any sensitivity displayed by the participants may be due to low level differences in novelty rather than any attribution of preferences. Thus, overall, I did not find any strong evidence for similarities between Eurasian jays' sensitivity to conspecifics' desires and adult humans' implicit sensitivity to others' desires.

This conclusion contrasts with the literature, which suggests convergence in cognition between humans and corvids (Emery & Clayton, 2004; Seed et al., 2009), and in which crows have been described as 'feathered apes' (Clayton, 2012). However, I argue that the social cognitive capabilities of corvids may have been over-interpreted relative to the available evidence. As I described in the previous section, there are many systemic reasons why a skewed view of the cognitive capabilities of a species may arise. However, the specific theoretical flexibility inherent in comparative cognition may exacerbate these issues. As I mentioned in Chapter 1, researchers have substantial freedom in what they consider to be evidence for theory of mind (Heyes, 2015).

The vagueness of verbal definitions of theory of mind (and other cognitive abilities) leads to disagreements over the interpretation of experimental results (Smaldino, 2017).

One specific issue is that researchers often conflate complex or flexible behaviours with the suggestion that this behaviour requires a conscious understanding of the situation (Richard W. Byrne & Bates, 2006). In fact, the 'cognitive' descriptor is often used synonymously with 'intelligent' or 'advanced' when describing behaviour. Heyes has labelled this stance as that of a 'cognition conservative', in contrast to a 'cognition liberal' stance, which defines cognition simply as adaptive information processing (Heyes, 2019). The disconnect between these two approaches arises when researchers design experiments based on a liberal definition of cognition (L-cognition), but then describe their results as indicative of conservative cognition (C-cognition). Simply handling information in an adaptive way (L-cognition), such as dropping a stone into a water-filled tube, does not imply that the individual has an understanding or conscious representation of the mechanics of the task (C-cognition) (Hennefield, Hwang, Weston, & Povinelli, 2018).

Similar semantic problems persist within the literature regarding the development of theory of mind in human infants. The classic violation of expectation experiment by Onishi & Baillargeon has been repeatedly cited as evidence that infants have an 'implicit understanding' of belief (Baillargeon et al., 2010; Onishi & Baillargeon, 2005). However, the authors never explain what it means to 'understand', leaving the language ambiguous (Heyes, 2012, 2019; Penn & Povinelli, 2007). This allows both higher and lower level interpretations, with early mind-reading accounts suggesting that infants have a full understanding of theory of mind and conceptual change accounts that suggest a qualitative distinction between infant and adult theory of mind (Low & Perner, 2012). Thus, flexibility in the defining of terms allows researchers and reviewers to interpret experimental studies to support whichever hypothesis they wish to promote (Buckner, 2013; Heyes, 2015).

The doublespeak involved in the discussion of cognition allows researchers to couch their conclusions in ambiguous language and avoid being accused of overselling their work; while simultaneously accessing high impact journals by using terms such as 'episodic memory', 'theory of mind' or 'teaching'. These labels are intrinsically linked in people's minds with C-cognition, and researchers title their papers in a manner that lends itself to C-cognitive interpretations. For example "Ravens parallel great apes in flexible planning for tool-use and bartering" (Kabadayi & Osvath, 2017), "...The western scrub-jay as a natural psychologist" (Clayton et al., 2007), "Chimpanzees really know what others can see in a competitive situation" (Bräuer, Call, & Tomasello, 2007). Titles of this type garner more positive attention and allow authors to publish

highly; however, the consequence of this practice is a superficially biased literature that encourages researchers (and funding bodies) to overestimate the cognitive abilities of non-human animals.

Moreover, this bias towards explaining behaviour in mentalistic terms may run even more deeply, affecting our ability to judge even our own behaviour. As I described in Chapter 1, early discussions of animals' psychological capabilities relied on the 'argument by analogy' (Hume, 1739; Romanes, 1883). A consequence of this approach is that there is a historic pattern of defining cognitive abilities, such as theory of mind, in an anthropocentric manner. This leads to underspecified, untestable definitions that themselves appeal to further subjective or under-defined concepts such as 'representation' or 'understanding' (Buckner, 2013; Heyes, 2015). As Buckner states, "these concepts are not even done baking in human psychology, and so deciding how to apply them to animals is especially fraught" (pg. 854) (Buckner, 2013). It is therefore likely that this lack of clarity in definition contributes to the very basis of disagreement between the sceptics and proponents of animal theory of mind.

The argument by analogy rests on the assumption that we are able to accurately assess our own psychological processes by introspection (Povinelli et al., 2000; Povinelli & Giambrone, 1999). However, there is increasing evidence that adult humans may not in fact be using the most strictly defined 'genuine' or 'full blown' theory of mind as frequently as generally assumed (Penn & Povinelli, 2013; Shettleworth, 2010). It has been suggested that rather than constantly invoking theory of mind, humans may use different strategies to predict behaviour such as personality traits, societal norms, or 'sub-mentalising' (Andrews, 2008, 2009; Heyes, 2014b). Although humans are able to retrospectively discuss both the behaviour of others and their own behaviour in a mentalistic manner, it is possible that such post-hoc reasoning belies the non-mentalistic nature of our own responses in the majority of social interactions.

Overestimating the extent to which we use theory of mind can have further consequences as it likely shapes the way we define and think about theory of mind in non-human animals. This extension of anthropocentrism has been described as the error of 'anthropofabulation', where an elevated impression of human cognition informs how we think about vaguely defined psychological abilities (Buckner, 2013). This causes problems when anthropofabulous definitions are combined with anthropomorphism such that an artificially inflated account of human cognition is then attributed to non-human animals on the basis of insufficient evidence.

Such a combination of errors might seem unlikely; however, there may be inherent biases within human psychology that encourage us to overestimate the prevalence of theory of mind in other species. Specifically, Povinelli et al.'s 'reinterpretation hypothesis' discusses a potential origin for

this over-attribution of mind-reading (Povinelli et al., 2000; Povinelli & Giambrone, 1999; Povinelli & Vonk, 2003). The premise of their hypothesis is that the derivation of theory of mind did not produce a suite of new behaviours, but rather allowed the reinterpretation of existing behaviours in mentalistic terms (Povinelli & Bering, 2002). However, this reinterpretation may overextend causing an anthropomorphic attribution of mental states, and by extension theory of mind, when it is not truly present. This over-mentalising may contribute to the easy assumption that animal's anecdotally engage in 'deception' or other mentalistic strategies (R. W. Byrne & Whiten, 1992; Povinelli & Vonk, 2003), as well as causing us to rationalise our own behaviour as mentalistic even when other explanations may be more plausible.

Finally, if this in-built over-mentalising is combined with the issues described above regarding an incentive system that favours positive results, as well as a trend to over-interpret findings, it seems highly likely that certain cognitive abilities are being attributed where evidence is insufficient. The lack of a clear definition of theory of mind at both the theoretical and the operational level has seriously hindered progress in the field. Agreement on whether non-human animals show any form of theory of mind cannot be reached until a valid definition has been agreed upon. However, the term 'theory of mind' may already be too burdened with existing connotations for this to be a realistic goal. While publications, funding and public interest continue to rely on vague, anthropofabulous definitions of theory of mind, constructive progress cannot be made on the question of mind-reading in non-human animals.

7.5 Concluding remarks

In this thesis I have addressed speculation within the field regarding the possibility of convergence in the social cognitive abilities of humans and non-human animals. In the first chapter I provided a comprehensive, critical overview of the existing evidence for theory of mind in non-human animals and the difficulties involved in successfully exploring this issue, which may be used to inform future research questions or strategies.

As the specific focus of this thesis, I asked whether implicit responses to desires in adult humans and Eurasian jays' responses to conspecifics' desires might involve a similar minimal theory of mind system. To address this question I designed a range of novel paradigms, which may be adapted in future for further exploration of theory of mind in both humans and non-human animals.

However, across my experiments with human adults, I did not find convincing evidence for implicit responses to desires in human adults. Similarly, my findings in Eurasian jays did not reflect the presence of a generalised minimal theory of mind, suggesting instead that the

response to satiety seen during beak-to-beak food sharing may be specific to the food sharing context. This therefore indicates that Eurasian jays' sensitivity to others' desires is more limited than previously believed and that claims of theory of mind in corvids are not fully justified by the existing evidence.

Contrastingly, in a further experiment I found that Eurasian jays appear to be able to respond flexibly to their partner's preferences in beak-to-beak food sharing, indicating that while they may not possess a generalised sensitivity to the desires of others, they nevertheless show cognitive flexibility in an ecologically valid context. While I did not find evidence for sensitivity to other's specific satiety during caching, future research could employ this choice-based paradigm to determine whether Eurasian jays might be sensitive to competitor's preferences in the caching context.

Finally, in this chapter I outlined the replication and reliability issues facing comparative cognition, stemming from a flawed incentive structure and ambiguous theoretical frameworks. I further highlight that the focus on high-level cognitive explanations to the exclusion of all else has likely led to the over-interpretation of available evidence regarding theory of mind. As such, I conclude that the lack of evidence for a shared minimal theory of mind mechanism found in this thesis may tentatively represent a true negative finding.

References

Allen, C. (2012). Transitive inference in animals: Reasoning or conditioned associations? In *Rational Animals?* (pp. 175–186). https://doi.org/10.1093/acprof:oso/9780198528272.003.0007

Allen, J. W. P. (2015). How to help: Can more active behavior measures help transcend the infant false-belief debate? *New Ideas in Psychology*, *39*, 63–72. https://doi.org/10.1016/j.newideapsych.2015.07.008

Amodio, P. (2019). *(Unpublished doctoral dissertation)*. University of Cambridge, UK.

Andersson, M., & Krebs, J. (1978). On the evolution of hoarding behaviour. *Animal Behaviour*, *26*(PART 3), 707–711. https://doi.org/10.1016/0003-3472(78)90137-9

Andrews, K. (2005). Chimpanzee Theory of Mind: Looking in All the Wrong Places? *Mind and Language*, *20*(5), 521–536. https://doi.org/10.1111/j.0268-1064.2005.00298.x

Andrews, K. (2008). It's in your nature: a pluralistic folk psychology. *Synthese*, *165*(1), 13–29. https://doi.org/10.1007/s11229-007-9230-5

Andrews, K. (2009). Understanding Norms Without a Theory of Mind. *Inquiry*, *52*(5), 433–448. https://doi.org/10.1080/00201740903302584

Apperly, I. A. (2010). Mindreaders: The Cognitive Basis of "Theory of Mind." In *Mindreaders: The Cognitive Basis of "Theory of Mind."* https://doi.org/10.4324/9780203833926

Apperly, I. A., Back, E., Samson, D., & France, L. (2008). The cost of thinking about false beliefs: evidence from adults' performance on a non-inferential theory of mind task. *Cognition*, *106*(3), 1093–1108. https://doi.org/10.1016/j.cognition.2007.05.005

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953–970. https://doi.org/10.1037/a0016923

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is Belief Reasoning Automatic? *Psychological Science*, *17*(10), 841–844. https://doi.org/10.1111/j.1467-9280.2006.01791.x

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of Adults Can Inform Accounts of Theory of Mind Development. *Developmental Psychology*, *45*(1), 190–201. https://doi.org/10.1037/a0014098

Astington, J. W., & Baird, J. A. (Eds.). (2005). *Why Language Matters for Theory of Mind*. https://doi.org/10.1093/acprof:oso/9780195159912.001.0001

Austin, G., Groppe, K., & Elsner, B. (2014). The reciprocal relationship between executive function and theory of mind in middle childhood: a 1-year longitudinal perspective. *Frontiers in Psychology*, *5*, 655. https://doi.org/10.3389/fpsyg.2014.00655

Back, E., & Apperly, I. A. (2010). Two sources of evidence on the non-automaticity of true and false belief ascription. *Cognition*, *115*(1), 54–70. https://doi.org/10.1016/j.cognition.2009.11.008

Baillargeon, R., Scott, R. M., & He, Z. (2010). False-belief understanding in infants. *Trends in Cognitive Sciences*, *14*(3), 110–118. https://doi.org/10.1016/j.tics.2009.12.006

Balleine, B. W., & Dickinson, A. (1998). The role of incentive learning in instrumental outcome revaluation by sensory-specific satiety. *Animal Learning & Behavior*, *26*(1), 46–59.

https://doi.org/10.3758/BF03199161

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of mind” ? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Beach, F. A. (1950). The Snark was a Boojum. *American Psychologist*, *5*(4), 115–124. https://doi.org/10.1037/h0056510

Beauchamp, G., & Fernández-Juricic, E. (2004). *Is there a relationship between forebrain size and group size in birds?* Retrieved from www.evolutionary-ecology.com

Bednekoff, P. A., & Balda, R. P. (1996). Observational spatial memory in Clark’s nutcrackers and Mexican jays. *Animal Behaviour*, *52*(4), 833–839. https://doi.org/10.1006/anbe.1996.0228

Beran, M. J. (2018). Replication and Pre-Registration in Comparative Psychology. *International Journal of Comparative Psychology*, *31*.

Bird, C. D., & Emery, N. J. (2008). Using video playback to investigate the social preferences of rooks, Corvus frugilegus. *Animal Behaviour*, *76*(3), 679–687. https://doi.org/10.1016/j.anbehav.2008.04.014

Bird, C. D., & Emery, N. J. (2009). Insightful problem solving and creative tool modification by captive nontool-using rooks. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10370–10375. https://doi.org/10.1073/pnas.0901008106

Bird, C. D., & Emery, N. J. (2010). Rooks perceive support relations similar to six-month-old babies. *Proceedings of the Royal Society B*, *277*(1678), 147–151. https://doi.org/10.1098/rspb.2009.1456

Biro, S., Verschoor, S., & Coenen, L. (2011). Evidence for a unitary goal concept in 12-month-old infants. *Developmental Science*, *14*(6), 1255–1260. https://doi.org/10.1111/j.1467-7687.2011.01042.x

Boissy, A., Manteuffel, G., Jensen, M. B., Moe, R. O., Spruijt, B., Keeling, L. J., … Aubert, A. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology & Behavior*, *92*(3), 375–397. https://doi.org/10.1016/J.PHYSBEH.2007.02.003

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, *10*(4), 439–448. https://doi.org/10.1007/s10071-007-0088-1

Bräuer, J., Call, J., & Tomasello, M. (2008). Chimpanzees do not take into account what others can hear in a competitive situation. *Animal Cognition*, *11*(1), 175–178. https://doi.org/10.1007/s10071-007-0097-0

Brecht, K. F., Ostojić, L., Legg, E. W., & Clayton, N. S. (2016). Working title: Can Eurasian Jays take into account the intention of their partner in a food-sharing context? Retrieved from https://osf.io/sg3d6/

Bryant, P. E., & Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature*, *232*(5311), 456–458. https://doi.org/10.1038/232456a0

Buckner, C. (2013). Morgan’s Canon, meet Hume’s Dictum: avoiding anthropofabulation in cross-species comparisons. *Biology & Philosophy*, *28*(5), 853–871. https://doi.org/10.1007/s10539-013-9376-0

Bugnyar, T. (2017). Apes perform like infants in false-belief tasks. *Learning and Behavior*, *45*(4), 325–326. https://doi.org/10.3758/s13420-017-0268-z

Bugnyar, T., & Heinrich, B. (2005). Ravens, Corvus corax, differentiate between knowledgeable and ignorant competitors. *Proceedings. Biological Sciences / The Royal Society*, *272*(1573), 1641–1646. https://doi.org/10.1098/rspb.2005.3144

Bull, R., Phillips, L. H., & Conway, C. A. (2008). The role of control functions in mentalizing: dual-task studies of theory of mind and executive function. *Cognition*, *107*(2), 663–672. https://doi.org/10.1016/j.cognition.2007.07.015

Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*(2), 287–314. https://doi.org/10.1016/J.COGNITION.2006.07.001

Burish, M. J., Kueh, H. Y., & Wang, S. S. H. (2004). Brain Architecture and Social Complexity in Modern and Ancient Birds. *Brain, Behavior and Evolution*, *63*(2), 107–124. https://doi.org/10.1159/000075674

Burnside, K., Ruel, A., Azar, N., & Poulin-Dubois, D. (2017). Implicit false belief across the lifespan: Non-replication of an anticipatory looking task. *Cognitive Development*. https://doi.org/10.1016/J.COGDEV.2017.08.006

Burt, D., & Peterson, A. (1993). Biology of Cooperative-Breeding Scrub Jays (Aphelocoma Coerulescens) of Oaxaca, Mexico. *The Auk: Ornithological Advances*, *110*(2), 207–214. https://doi.org/10.1093/auk/110.2.207

Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., & Tomasello, M. (2017). Great apes distinguish true from false beliefs in an interactive helping task. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0173793

Buttelmann, D., Call, J., & Tomasello, M. (2009). Do great apes use emotional expressions to infer desires? *Developmental Science*, *12*(5), 688–698. https://doi.org/10.1111/j.1467-7687.2008.00802.x

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2007). Enculturated chimpanzees imitate rationally. *Developmental Science*, *10*(4), F31-8. https://doi.org/10.1111/j.1467-7687.2007.00630.x

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337–342. https://doi.org/10.1016/j.cognition.2009.05.006

Buttelmann, D., Over, H., Carpenter, M., & Tomasello, M. (2014). Eighteen-month-olds understand false beliefs in an unexpected-contents task. *Journal of Experimental Child Psychology*, *119*, 120–126. https://doi.org/10.1016/j.jecp.2013.10.002

Buttelmann, F., Suhrke, J., & Buttelmann, D. (2015). What you get is what you believe: Eighteen-month-olds demonstrate belief understanding in an unexpected-identity task. *Journal of Experimental Child Psychology*, *131*, 94–103. https://doi.org/10.1016/j.jecp.2014.11.009

Butterfill, S. A., & Apperly, I. A. (2013). How to Construct a Minimal Theory of Mind. *Mind & Language*, *28*(5), 606–637. https://doi.org/10.1111/mila.12036

Butterfill, S. A., & Apperly, I. A. (2016). Is Goal Ascription Possible in Minimal Mindreading? *Psychological Review*, *123*(2), 228–233. https://doi.org/10.1037/rev0000022

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews*

Neuroscience, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Byrne, R. W., & Whiten, A. (1992). Cognitive Evolution in Primates: Evidence from Tactical Deception. *Man*, *27*(3), 609. https://doi.org/10.2307/2803931

Byrne, Richard W., & Bates, L. A. (2006). Why are animals cognitive? *Current Biology*, *16*(12), R445–R448. https://doi.org/10.1016/J.CUB.2006.05.040

Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). "Unwilling" versus "unable": chimpanzees' understanding of human intentional action. *Developmental Science*, *7*(4), 488–498. https://doi.org/10.1111/j.1467-7687.2004.00368.x

Call, J., & Tomasello, M. (1998). Distinguishing intentional from accidental actions in orangutans (Pongo pygmaeus), chimpanzees (Pan troglodytes), and human children (Homo sapiens). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *112*(2), 192–206. https://doi.org/10.1037/0735-7036.112.2.192

Call, J., & Tomasello, M. (1999). A Nonverbal False Belief Task: The Performance of Children and Great Apes. *Child Development*, *70*(2), 381–395. https://doi.org/10.1111/1467-8624.00028

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Carruthers, P. (2013). Mindreading in Infancy. *Mind & Language*, *28*(2), 141–172. https://doi.org/10.1111/mila.12014

Cheke, L. G., Bird, C. D., & Clayton, N. S. (2011). Tool-use and instrumental learning in the Eurasian jay (Garrulus glandarius). *Animal Cognition*, *14*(3), 441–455. https://doi.org/10.1007/s10071-011-0379-4

Clary, D., & Kelly, D. M. (2011). Cache protection strategies of a non-social food-caching corvid, Clark's nutcracker (Nucifraga columbiana). *Animal Cognition*, *14*(5), 735–744. https://doi.org/10.1007/s10071-011-0408-3

Clayton, N. S. (2012). Corvid cognition: Feathered apes. *Nature*, *484*(7395), 453–454. https://doi.org/10.1038/484453a

Clayton, N. S., Dally, J. M., & Emery, N. J. (2007). Social cognition by food-caching corvids. The western scrub-jay as a natural psychologist. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1480), 507–522. https://doi.org/10.1098/rstb.2006.1992

Clayton, N. S., Dally, J. M., Gilbert, J., & Dickinson, A. (2005). Food Caching by Western Scrub-Jays (Aphelocoma californica) Is Sensitive to the Conditions at Recovery. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(2), 115–124. https://doi.org/10.1037/0097-7403.31.2.115

Clayton, N. S., & Emery, N. J. (2007, August 21). The social life of corvids. *Current Biology*, Vol. 17. https://doi.org/10.1016/j.cub.2007.05.070

Cohen, J. (1994). The Earth is Round ($p < .05$). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Csibra, G., & Gergely, G. (1998). The teleological origins of mentalistic action explanations: A

developmental hypothesis. *Developmental Science*, *1*(2), 255–259. https://doi.org/10.1111/1467-7687.00039

Dally, J. M., Emery, N. J., & Clayton, N. S. (2004a). Cache protection strategies by western scrub-jays (Aphelocoma californica): hiding food in the shade. *Proceedings. Biological Sciences / The Royal Society*, *271 Suppl*(Suppl_6), S387-90. https://doi.org/10.1098/rsbl.2004.0190

Dally, J. M., Emery, N. J., & Clayton, N. S. (2004b). Cache protection strategies by western scrub–jays (Aphelocoma californica): hiding food in the shade. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(suppl_6). https://doi.org/10.1098/rsbl.2004.0190

Dally, J. M., Emery, N. J., & Clayton, N. S. (2005). Cache protection strategies by western scrub-jays, Aphelocoma californica: implications for social cognition. *Animal Behaviour*, *70*(6), 1251–1263. https://doi.org/10.1016/j.anbehav.2005.02.009

Dally, J. M., Emery, N. J., & Clayton, N. S. (2006). Food-caching western scrub-jays keep track of who was watching when. *Science*, *312*(5780), 1662–1665. https://doi.org/10.1126/science.1126539

Dally, J. M., Emery, N. J., & Clayton, N. S. (2010). Avian Theory of Mind and counter espionage by food-caching western scrub-jays (Aphelocoma californica). *European Journal of Developmental Psychology*, *7*(1), 17–37. https://doi.org/10.1080/17405620802571711

Davidson, G., Miller, R., Loissel, E., Cheke, L. G., & Clayton, N. S. (2017). The development of support intuitions and object causality in juvenile Eurasian jays (Garrulus glandarius). *Scientific Reports*, *7*(January), 40062. https://doi.org/10.1038/srep40062

De Kort, S. R., & Clayton, N. S. (2006). An evolutionary perspective on caching by corvids. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1585), 417–423. https://doi.org/10.1098/rspb.2005.3350

de Kort, S. R., Correia, S. P. C., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). The control of food-caching behavior by Western scrub-jays (Aphelocoma californica). *Journal of Experimental Psychology: Animal Behavior Processes*, *33*(4), 361–370. https://doi.org/10.1037/0097-7403.33.4.361

De Kort, S. R., Emery, N. J., & Clayton, N. S. (2012). *Corvid CachingThe Role of Cognition*. https://doi.org/10.1093/OXFORDHB/9780195392661.013.0021

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*. https://doi.org/10.1016/J.COGDEV.2018.01.001

Dunbar, R. I. M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, *22*(6), 469–493. https://doi.org/10.1016/0047-2484(92)90081-J

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews*, *6*(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Emery, N. J., & Clayton, N. S. (2001). Effects of experience and social context on prospective caching strategies by scrub jays. *Nature*, *414*(6862), 443–446. https://doi.org/10.1038/35106560

Emery, N. J., & Clayton, N. S. (2004). The Mentality of Crows: Convergent Evolution of Intelligence in

Corvids and Apes. *Science*, *306*(5703), 1903–1907. https://doi.org/10.1126/science.1098410

Emery, N. J., Clayton, N. S., & Frith, C. D. (2007). Introduction. Social intelligence: From brain to culture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 485–488. https://doi.org/10.1098/rstb.2006.2022

Emery, N. J., Dally, J. M., & Clayton, N. S. (2004). Western scrub-jays (Aphelocoma californica) use cognitive strategies to protect their caches from thieving conspecifics. *Animal Cognition*, *7*(1), 37–43. https://doi.org/10.1007/s10071-003-0178-7

Emery, N. J., Seed, A. M., von Bayern, A. M. P., & Clayton, N. S. (2007). Cognitive adaptations of social bonding in birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 489–505. https://doi.org/10.1098/rstb.2006.1991

Farrar, B., Boeckle, M., & Clayton, N. (2020). Replications in Comparative Cognition: What Should We Expect and How Can We Improve? *Animal Behavior and Cognition*, *7*(1), 1–22. https://doi.org/10.26451/abc.07.01.02.2020

Ferguson, C. J., & Heene, M. (2012). A Vast Graveyard of Undead Theories. *Perspectives on Psychological Science*, *7*(6), 555–561. https://doi.org/10.1177/1745691612459059

Fiedler, K., & Schwarz, N. (2016). Questionable Research Practices Revisited. *Social Psychological and Personality Science*, *7*(1), 45–52. https://doi.org/10.1177/1948550615612150

Flombaum, J. I., & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology : CB*, *15*(5), 447–452. https://doi.org/10.1016/j.cub.2004.12.076

Frank, M. J., Rudy, J. W., Levy, W. B., & O'Reilly, R. C. (2005). When logic fails: Implicit transitive inference in humans. *Memory and Cognition*. https://doi.org/10.3758/BF03195340

Frank, M. J., Rudy, J. W., & O'Reilly, R. C. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus. II. A computational analysis. *Hippocampus*, *13*(3), 341–354. https://doi.org/10.1002/hipo.10084

Garnham, W. A., & Perner, J. (2001). Actions really do speak louder than words-but only implicitly: Young children's understanding of false belief in action. *British Journal of Developmental Psychology*, *19*(3), 413–432. https://doi.org/10.1348/026151001166182

Gazes, R. P., Hampton, R. R., & Lourenco, S. F. (2017). Transitive inference of social dominance by human infants. *Developmental Science*, *20*(2), e12367. https://doi.org/10.1111/desc.12367

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Gelman, A., & Stern, H. (2006). The Difference Between "Significant" and "Not Significant" is not Itself Statistically Significant. *The American Statistician*, *60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences*, *7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

German, T. P., & Hehman, J. A. (2006). Representational and executive selection resources in 'theory of mind': Evidence from compromised belief-desire reasoning in old age. *Cognition*, *101*(1), 129–152.

https://doi.org/10.1016/j.cognition.2005.05.007

Goodwin, D. (1951). Some apsects of the behaviour of the jay Garrulus Glandarius. *Ibis*, *93*(4), 602–625. https://doi.org/10.1111/j.1474-919X.1951.tb05462.x

Goodwin, D. (1956). Further observations on the behaviour of the jay (Garrulus glandarius). *Ibis*, *98*(2), 186–219. https://doi.org/10.1111/j.1474-919X.1956.tb03040.x

Greene, A. J., Spellman, B. A., Dusek, J. A., Eichenbaum, H. B., & Levy, W. B. (2001). Relational learning with and without awareness: Transitive inference using nonverbal stimuli in humans. *Memory and Cognition*. https://doi.org/10.3758/BF03196418

Griskevicius, V., Tybur, J. M., Sundie, J. M., Cialdini, R. B., Miller, G. F., & Kenrick, D. T. (2007). Blatant benevolence and conspicuous consumption: When romantic motives elicit strategic costly signals. *Journal of Personality and Social Psychology*, *93*(1), 85–102. https://doi.org/10.1037/0022-3514.93.1.85

Grodzinski, U., & Clayton, N. S. (2010). Problems faced by food-caching corvids and the evolution of cognitive solutions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1542), 977–987. https://doi.org/10.1098/rstb.2009.0210

Grodzinski, U., Watanabe, A., & Clayton, N. S. (2012). Peep to pilfer: what scrub-jays like to watch when observing others. *Animal Behaviour*, *83*(5), 1253–1260. https://doi.org/10.1016/j.anbehav.2012.02.018

Halina, M. (2015). *There Is No Special Problem of Mindreading in Nonhuman Animals*. Retrieved from http://people.ds.cam.ac.uk/mh801/halina-2015.pdf

Hare, B. (2001). Can competitive paradigms increase the validity of experiments on primate social cognition? *Animal Cognition*, *4*(3–4), 269–280. https://doi.org/10.1007/s100710100084

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771–785. https://doi.org/10.1006/anbe.1999.1377

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*(1), 139–151. https://doi.org/10.1006/anbe.2000.1518

Hare, B., Call, J., & Tomasello, M. (2006). Chimpanzees deceive a human competitor by hiding. *Cognition*, *101*(3), 495–514. https://doi.org/10.1016/j.cognition.2005.01.011

Hare, B., & Tomasello, M. (2004). Chimpanzees are more skilful in competitive than in cooperative cognitive tasks. *Animal Behaviour*, *68*(3), 571–581. https://doi.org/10.1016/J.ANBEHAV.2003.11.011

Havermans, R. C., Janssen, T., Giesen, J. C. A. H., Roefs, A., & Jansen, A. (2009). Food liking, food wanting, and sensory-specific satiety. *Appetite*, *52*(1), 222–225. https://doi.org/10.1016/j.appet.2008.09.020

Henderson, A. M. E., & Woodward, A. L. (2012). Nine-month-old infants generalize object labels, but not object preferences across individuals. *Developmental Science*, *15*(5), 641–652. https://doi.org/10.1111/j.1467-7687.2012.01157.x

Hennefield, L., Hwang, H. G., Weston, S. J., & Povinelli, D. J. (2018). Meta-analytic techniques reveal that corvid causal reasoning in the Aesop's Fable paradigm is driven by trial-and-error learning. *Animal Cognition*, *21*(6), 735–748. https://doi.org/10.1007/s10071-018-1206-y

Hernik, M., & Southgate, V. (2012). Nine-months-old infants do not need to know what the agent prefers in order to reason about its goals: on the role of preference and persistence in infants' goal-attribution. *Developmental Science*, *15*(5), 714–722. https://doi.org/10.1111/j.1467-7687.2012.01151.x

Hetherington, M. M., & Rolls, B. J. (1996). Sensory-specific satiety: Theoretical frameworks and central characteristics. In *Why we eat what we eat: The psychology of eating.* https://doi.org/10.1037/10291-010

Heyes, C. (1998). Theory of mind in nonhuman primates. *The Behavioral and Brain Sciences*, *21*(1), 101–114.

Heyes, C. (2012). Simple minds: a qualified defence of associative learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1603), 2695–2703. https://doi.org/10.1098/rstb.2012.0217

Heyes, C. (2014a). False belief in infancy: a fresh look. *Developmental Science*, *17*(5), 647–659. https://doi.org/10.1111/desc.12148

Heyes, C. (2014b). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *9*(2), 131–143. https://doi.org/10.1177/1745691613518076

Heyes, C. (2015). Animal mindreading: what's the problem? *Psychonomic Bulletin & Review*, *22*(2), 313–327. https://doi.org/10.3758/s13423-014-0704-4

Heyes, C. (2017). Apes Submentalise. *Trends in Cognitive Sciences*, *21*(1), 1–2. https://doi.org/10.1016/j.tics.2016.11.006

Heyes, C. (2019). What is cognition? *Current Biology*, *29*(13), R611. https://doi.org/10.1016/j.cub.2019.05.044

Horn, L., Scheer, C., Bugnyar, T., & Massen, J. J. M. (2016). Proactive prosociality in a cooperatively breeding corvid, the azure-winged magpie (*Cyanopica cyana*). *Biology Letters*, *12*(10), 20160649. https://doi.org/10.1098/rsbl.2016.0649

Horowitz, A. (2011). Theory of mind in dogs? Examining method and concept. *Learning and Behavior*, *39*(4), 314–317. https://doi.org/10.3758/s13420-011-0041-7

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*(2), 233–253. https://doi.org/10.1111/j.2044-835X.1998.tb00921.x

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology*, *43*(6), 1447–1459. https://doi.org/10.1037/0012-1649.43.6.1447

Hume, D. (1739). *A Treatise of Human Nature*. https://doi.org/10.2307/2216614

Humphrey, N. K. (1976). The social function of intellect. In P. P. G. . Bateson & R. A. Hinde (Eds.), *Growing points in ethology*. Cambridge University Press.

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling. *Psychological Science*, *23*(5), 524–532. https://doi.org/10.1177/0956797611430953

Johnson, S. C., Ok, S.-J., & Luo, Y. (2007). The attribution of attention: 9-month-olds? interpretation of gaze as goal-directed action. *Developmental Science*, *10*(5), 530–537. https://doi.org/10.1111/j.1467-7687.2007.00606.x

Kabadayi, C., & Osvath, M. (2017). Ravens parallel great apes in flexible planning for tool-use and bartering. *Science (New York, N.Y.)*, *357*(6347), 202–204. https://doi.org/10.1126/science.aam8138

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, *109*(2), 224–234. https://doi.org/10.1016/j.cognition.2008.08.010

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2015). The goggles experiment: can chimpanzees use self-experience to infer what a competitor can see? *Animal Behaviour*, *105*, 211–221. https://doi.org/10.1016/J.ANBEHAV.2015.04.028

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Koops, K., Furuichi, T., & Hashimoto, C. (2015). Chimpanzees and bonobos differ in intrinsic motivation for tool use. *Scientific Reports*, *5*(1), 11356. https://doi.org/10.1038/srep11356

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830–1834. https://doi.org/10.1126/science.1190792

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, *12*(4), 521–535. https://doi.org/10.1111/j.1467-7687.2008.00793.x

Krebs, J. R. (1990). Food-storing birds: adaptive specialization in brain and behaviour? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *329*(1253), 153–160. https://doi.org/10.1098/rstb.1990.0160

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2017). A test of the submentalizing hypothesis: Apes' performance in a false belief task inanimate control. *Communicative & Integrative Biology*, *10*(4), e1343771. https://doi.org/10.1080/19420889.2017.1343771

Kulke, L., & Rakoczy, H. (2018). Implicit Theory of Mind - An overview of current replications and non-replications. *Data in Brief*, *16*, 101–104. https://doi.org/10.1016/j.dib.2017.11.016

Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2017). Implicit Theory of Mind across the life span - Anticipatory looking data. *Data in Brief*, *15*, 712–719. https://doi.org/10.1016/j.dib.2017.10.021

Kulke, L., Reiß, M., Krist, H., & Rakoczy, H. (2018). How robust are anticipatory looking measures of Theory of Mind? Replication attempts across the life span. *Cognitive Development*, *46*, 97–111. https://doi.org/10.1016/J.COGDEV.2017.09.001

Kulke, L., von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is Implicit Theory of Mind a Real and Robust Phenomenon? Results From a Systematic Replication Study. *Psychological Science*, *29*(6), 888–900. https://doi.org/10.1177/0956797617747090

Lack, D. (1940). Courtship feeding in birds. *The Auk*, *57*, 169–178.

Lazareva, O. F. (2012). Transitive Inference in Nonhuman Animals. In *The Oxford Handbook of Comparative Cognition*. https://doi.org/10.1093/oxfordhb/9780195392661.013.0036

LeBel, E. P., & Peters, K. R. (2011). *Fearing the Future of Empirical Psychology: Bem's (2011) Evidence of Psi as a Case Study of Deficiencies in Modal Research Practice*. https://doi.org/10.1037/a0025172

Legg, E. W. (2014a). *A comparative approach to testing mental state attribution (Doctoral dissertation)*. University of Cambridge, UK.

Legg, E. W. (2014b). *Food sharing choice study (Unpublished raw data)*.

Legg, E. W., & Clayton, N. S. (2014). Eurasian jays (Garrulus glandarius) conceal caches from onlookers. *Animal Cognition*, *17*(5), 1223–1226. https://doi.org/10.1007/s10071-014-0743-2

Legg, E. W., Ostojić, L., & Clayton, N. S. (2016). Caching at a distance: a cache protection strategy in Eurasian jays. *Animal Cognition*. https://doi.org/10.1007/s10071-016-0972-7

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556. https://doi.org/10.1016/j.jesp.2009.12.019

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: State of the art. *British Journal of Developmental Psychology*, *30*(1), 1–13. https://doi.org/10.1111/j.2044-835X.2011.02074.x

Low, J., & Watts, J. (2013). Attributing False Beliefs About Object Identity Reveals a Signature Blind Spot in Humans' Efficient Mind-Reading System. *Psychological Science*, *24*(3), 305–311. https://doi.org/10.1177/0956797612451469

Lucas, J. R., Brodin, A., De Kort, S. R., & Clayton, N. S. (2004). Does hippocampal size correlate with the degree of caching specialization? *Proceedings of the Royal Society B: Biological Sciences*, *271*(1556), 2423–2429. https://doi.org/10.1098/rspb.2004.2912

Luo, Y. (2011). Three-month-old infants attribute goals to a non-human agent. *Developmental Science*, *14*(2), 453–460. https://doi.org/10.1111/j.1467-7687.2010.00995.x

Luo, Y., & Baillargeon, R. (2005). Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science*, *16*(8), 601–608. https://doi.org/10.1111/j.1467-9280.2005.01582.x

Luo, Y., & Baillargeon, R. (2007). Do 12.5-month-old infants consider what objects others can see when interpreting their actions? *Cognition*, *105*(3), 489–512. https://doi.org/10.1016/j.cognition.2006.10.007

Luo, Y., & Choi, Y. (2012). Infants attribute to agents goals and dispositions. *Developmental Science*, *15*(5), 727–728. https://doi.org/10.1111/j.1467-7687.2012.01163.x

Lurz, R. (2009). If chimpanzees are mindreaders, could behavioral science tell? Toward a solution of the logical problem. *Philosophical Psychology*, *22*(3), 305–328.

https://doi.org/10.1080/09515080902970673

M. Macphail, E., & J. Bolhuis, J. (2001). The evolution of intelligence: adaptive specializations versus general process. *Biological Reviews of the Cambridge Philosophical Society*, *76*(3), 341–364. https://doi.org/10.1017/s146479310100570x

Markovits, H., & Dumas, C. (1992). Can pigeons really make transitive inferences? *Journal of Experimental Psychology: Animal Behavior Processes*, *18*(3), 311–312. https://doi.org/10.1037/0097-7403.18.3.311

Martin, G. N., & Clarke, R. M. (2017). Are Psychology Journals Anti-replication? A Snapshot of Editorial Practices. *Frontiers in Psychology*, *8*, 523. https://doi.org/10.3389/fpsyg.2017.00523

Martinho, A., Burns, Z. T., von Bayern, A. M. P., & Kacelnik, A. (2014). Monocular Tool Control, Eye Dominance, and Laterality in New Caledonian Crows. *Current Biology*, *24*(24), 2930–2934. https://doi.org/10.1016/J.CUB.2014.10.035

Mascaro, O., & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1113194109

Mascaro, O., & Csibra, G. (2014). Human Infants' Learning of Social Structures. *Psychological Science*, *25*(1), 250–255. https://doi.org/10.1177/0956797613500509

McKinnon, M. C., & Moscovitch, M. (2007). Domain-general contributions to social reasoning: Theory of mind and deontic reasoning re-explored. *Cognition*, *102*(2), 179–218. https://doi.org/10.1016/J.COGNITION.2005.12.011

Meehl, P. E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science*, *34*(2), 103–115. https://doi.org/10.1086/288135

Melis, A. P., Call, J., & Tomasello, M. (2006). Chimpanzees (Pan troglodytes) conceal visual and auditory information from others. *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *120*(2), 154–162. https://doi.org/10.1037/0735-7036.120.2.154

Melis, A. P., Warneken, F., Jensen, K., Schneider, A.-C., Call, J., & Tomasello, M. (2011). Chimpanzees help conspecifics obtain food and non-food items. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1710), 1405–1413. https://doi.org/10.1098/rspb.2010.1735

Meltzoff, A. N., & Brooks, R. (2008). Self-Experience as a Mechanism for Learning About Others: A Training Study in Social Cognition. *Developmental Psychology*, *44*(5), 1257–1265. https://doi.org/10.1037/a0012888

Mou, Y., Province, J. M., & Luo, Y. (2014). Can infants make transitive inferences? *Cognitive Psychology*, *68*, 98–112. https://doi.org/10.1016/J.COGPSYCH.2013.11.003

Munafó, M. R. (2009, September 1). Repliciability and replicability of genetic association studies. *Addiction*, Vol. 104, pp. 1439–1440. https://doi.org/10.1111/j.1360-0443.2009.02662.x

Nichols, S., & Stich, S. P. (2003). *Mindreading: an integrated account of pretence, self-awareness, and understanding other minds.* https://doi.org/10.1093/0198236107.003.0003

Nissen, S. B., Magidson, T., Gross, K., & Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *ELife*, *5*. https://doi.org/10.7554/eLife.21451

Nottebohm, F. (1981). A brain for all seasons: cyclical anatomical changes in song control nuclei of the canary brain. *Science*, *214*(4527), 1368–1370. https://doi.org/10.1126/science.7313697

O'Connell, S., & Dunbar, R. I. M. (2003). A Test for Comprehension of False Belief in Chimpanzees. *Evolution and Cognition*, *9*, 131–139. https://doi.org/10.1007/s10071-004-0231-1

Onishi, K. H., & Baillargeon, R. (2005). Do 15-Month-Old Infants Understand False Beliefs? *Science*, *308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Ostojić, L., Cheke, L. G., Shaw, R. C., Legg, E. W., & Clayton, N. S. (2016). Desire-state attribution: Benefits of a novel paradigm using the food-sharing behavior of eurasian jays (Garrulus glandarius). *Communicative and Integrative Biology*, *9*(2), 1–3. https://doi.org/10.1080/19420889.2015.1134065

Ostojić, L., Legg, E. W., Brecht, K. F., Lange, F., Deininger, C., Mendl, M., & Clayton, N. S. (2017). Current desires of conspecific observers affect cache-protection strategies in California scrub-jays and Eurasian jays. *Current Biology*, *27*(2), R51–R53. https://doi.org/10.1016/J.CUB.2016.11.020

Ostojić, L., Legg, E. W., Dits, A., Williams, N., Brecht, K. F., Mendl, M., & Clayton, N. S. (2016). Experimenter Expectancy Bias Does Not Explain Eurasian Jays' (Garrulus glandarius) Performance in a Desire-State Attribution Task. *Journal of Comparative Psychology*. https://doi.org/10.1037/com0000043

Ostojić, L., Legg, E. W., Mendl, M., & Clayton, N. S. (2014). *Food sharing inference study (Unpublished raw data)*.

Ostojić, L., Legg, E. W., Shaw, R. C., Cheke, L. G., Mendl, M., & Clayton, N. S. (2014). Can male Eurasian jays disengage from their own current desire to feed the female what she wants? *Biology Letters*, *10*(3), 20140042. https://doi.org/10.1098/rsbl.2014.0042

Ostojić, L., Shaw, R. C., Cheke, L. G., & Clayton, N. S. (2013). Evidence suggesting that desire-state attribution may govern food sharing in Eurasian jays. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(10), 4123–4128. https://doi.org/10.1073/pnas.1209926110

Palan, S., & Schitter, C. (2018). Prolific.ac—A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, *17*, 22–27. https://doi.org/10.1016/J.JBEF.2017.12.004

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a "theory of mind". *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *362*(1480), 731–744. https://doi.org/10.1098/rstb.2006.2023

Penn, D. C., & Povinelli, D. J. (2013). The Comparative Delusion: The "Behavioristic/Mentalistic" Dichotomy in Comparative Theory of Mind Research. In *Agency and Joint Attention* (pp. 62–81). Oxford University Press.

Perner, J., & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences*, *16*(10), 519–525. https://doi.org/10.1016/j.tics.2012.08.004

Perner, J., & Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science*, *308*(5719), 214–216. https://doi.org/10.1126/science.1111656

Phillips, A. T., Wellman, H. M., & Spelke, E. S. (2002). Infants' ability to connect gaze and emotional expression to intentional action. *Cognition*, *85*(1), 53–78. https://doi.org/10.1016/S0010-0277(02)00073-2

Phillips, J., Ong, D. C., Surtees, A. D. R., Xin, Y., Williams, S., Saxe, R., & Frank, M. C. (2015). A Second Look at Automatic Theory of Mind. *Psychological Science*, *26*(9), 1353–1367. https://doi.org/10.1177/0956797614558717

Poulin-Dubois, D., Rakoczy, H., Burnside, K., Crivello, C., Dörrenberg, S., Edwards, K., … Ruffman, T. (2018). Do infants understand false beliefs? We don't know yet – A commentary on Baillargeon, Buttelmann and Southgate's commentary. *Cognitive Development*, *48*, 302–315. https://doi.org/10.1016/J.COGDEV.2018.09.005

Povinelli, D. J., & Bering, J. M. (2002). The Mentality of Apes Revisited. *Current Directions in Psychological Science*, *11*(4), 115–119. https://doi.org/10.1111/1467-8721.00181

Povinelli, D. J., Bering, J. M., & Giambrone, S. (2000). Toward a Science of Other Minds: Escaping the Argument by Analogy. *Cognitive Science*, *24*(3), 509–541. https://doi.org/10.1207/s15516709cog2403_7

Povinelli, D. J., & Eddy, T. J. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, *61*(3), i–189. https://doi.org/10.2307/1166159

Povinelli, D. J., & Giambrone, S. (1999). *Inferring Other Minds: Failure of the Argument by Analogy*. https://doi.org/10.5840/philtopics199927120

Povinelli, D. J., Perilloux, H. K., Reaux, J. E., & Bierschwale, D. T. (1998). Young and juvenile chimpanzees' (Pan troglodytes) reactions to intentional versus accidental and inadvertent actions. *Behavioural Processes*, *42*(2–3), 205–218. https://doi.org/10.1016/S0376-6357(97)00077-6

Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends in Cognitive Sciences*, *7*(4), 157–160. https://doi.org/10.1016/S1364-6613(03)00053-6

Povinelli, D. J., & Vonk, J. (2004). We Don't Need a Microscope to Explore the Chimpanzee's Mind. *Mind and Language*, *19*(1), 1–28. https://doi.org/10.1111/j.1468-0017.2004.00244.x

Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development*, *46*, 40–50. https://doi.org/10.1016/J.COGDEV.2017.10.004

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(04), 515. https://doi.org/10.1017/S0140525X00076512

Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, *10*(9), 712–712. https://doi.org/10.1038/nrd3439-c1

Raby, C. R., Alexis, D. M., Dickinson, A., & Clayton, N. S. (2007). Planning for the future by western scrub-jays. *Nature*, *445*(7130), 919–921. https://doi.org/10.1038/nature05575

Rakoczy, H., Bergfeld, D., Schwarz, I., & Fizke, E. (2015). Explicit Theory of Mind Is Even More Unified Than Previously Assumed: Belief Ascription and Understanding Aspectuality Emerge Together in

Development. *Child Development*, *86*(2), 486–502. https://doi.org/10.1111/cdev.12311

Range, F., & Huber, L. (2007). Attention in common marmosets: implications for social-learning experiments. *Animal Behaviour*, *73*(6), 1033–1041. https://doi.org/10.1016/J.ANBEHAV.2006.07.015

Reaux, J. E., Theall, L. A., & Povinelli, D. J. (1999). A Longitudinal Investigation of Chimpanzees' Understanding of Visual Perception. *Child Development*, *70*(2), 275–290. https://doi.org/10.1111/1467-8624.00021

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: evidence from 14- and 18-month-olds. *Developmental Psychology*, *33*(1), 12–21. https://doi.org/10.1037/0012-1649.33.1.12

Riters, L. V. (2012). The role of motivation and reward neural systems in vocal communication in songbirds. *Frontiers in Neuroendocrinology*, *33*(2), 194–209. https://doi.org/10.1016/J.YFRNE.2012.04.002

Roberts, W. A., & Phelps, M. T. (1994). Transitive Inference in Rats: A Test of the Spatial Coding Hypothesis. *Psychological Science*, *5*(6), 368–374. https://doi.org/10.1111/j.1467-9280.1994.tb00287.x

Romanes, G. J. (1883). *Mental evolution in animals*. https://doi.org/10.1017/CBO9781139103664

Rosati, A. G. (2017, September 1). Foraging Cognition: Reviving the Ecological Intelligence Hypothesis. *Trends in Cognitive Sciences*, Vol. 21, pp. 691–702. https://doi.org/10.1016/j.tics.2017.05.011

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Ruffman, T., Taumoepeau, M., & Perkins, C. (2012). Statistical learning as a basis for social understanding in children. *British Journal of Developmental Psychology*, *30*(1), 87–104. https://doi.org/10.1111/j.2044-835X.2011.02045.x

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The Development of Executive Functioning and Theory of Mind. A Comparison of Chinese and U.S. Preschoolers. *Psychological Science*, *17*(1), 74–81. https://doi.org/10.1111/j.1467-9280.2005.01667.x

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Scott, S. E. B. (2010). *Seeing It Their Way: Evidence for Rapid and Involuntary Computation of What Other People See*. https://doi.org/10.1037/a0018729

Santiesteban, I., Catmur, C., Hopkins, S. C., Bird, G., & Heyes, C. (2014). Avatars and arrows: Implicit mentalizing or domain-general processing? *Journal of Experimental Psychology: Human Perception and Performance*, *40*(3), 929–937. https://doi.org/10.1037/a0035175

Santos, L. R., Nissen, A. G., & Ferrugia, J. A. (2006). Rhesus monkeys, Macaca mulatta, know what others can and cannot hear. *Animal Behaviour*, *71*(5), 1175–1181. https://doi.org/10.1016/j.anbehav.2005.10.007

Scheid, C., Range, F., & Bugnyar, T. (2007). When, what, and whom to watch? Quantifying attention in ravens (Corvus corax) and jackdaws (Corvus monedula). *Journal of Comparative Psychology*. https://doi.org/10.1037/0735-7036.121.4.380

Schuwerk, T., Vuori, M., & Sodian, B. (2015). Implicit and explicit Theory of Mind reasoning in autism spectrum disorders: The impact of experience. *Autism*, *19*(4), 459–468. https://doi.org/10.1177/1362361314526004

Schwab, C., Swoboda, R., Kotrschal, K., & Bugnyar, T. (2012). Recipients Affect Prosocial and Altruistic Choices in Jackdaws, Corvus monedula. *PLoS ONE*, *7*(4), e34922. https://doi.org/10.1371/journal.pone.0034922

Scott, R. M., & Baillargeon, R. (2014). How fresh a look? A reply to Heyes. *Developmental Science*, *17*(5), 660–664. https://doi.org/10.1111/desc.12173

Scott, R. M., & Baillargeon, R. (2017). Early False-Belief Understanding. *Trends in Cognitive Sciences*, *21*(4), 237–249. https://doi.org/10.1016/j.tics.2017.01.012

Seed, A. M., Emery, N. J., & Clayton, N. S. (2009). Intelligence in Corvids and Apes: A Case of Convergent Evolution? *Ethology*, *115*(5), 401–420. https://doi.org/10.1111/j.1439-0310.2009.01644.x

Senju, A., Southgate, V., Snape, C., Leonard, M., & Csibra, G. (2011). Do 18-month-olds really attribute mental states to others? A critical test. *Psychological Science*, *22*(7), 878–880. https://doi.org/10.1177/0956797611411584

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science*, *219*(August), 883–885. https://doi.org/10.1126/science.1176170

Shaw, R. C., & Clayton, N. S. (2012). Eurasian jays, Garrulus glandarius, flexibly switch caching and pilfering tactics in response to social context. *Animal Behaviour*, *84*(5), 1191–1200. https://doi.org/10.1016/j.anbehav.2012.08.023

Shaw, R. C., & Clayton, N. S. (2013). Careful cachers and prying pilferers: Eurasian jays (Garrulus glandarius) limit auditory information available to competitors. *Proceedings. Biological Sciences / The Royal Society*, *280*(1752), 20122238. https://doi.org/10.1098/rspb.2012.2238

Shettleworth, S. J. (1990). Spatial memory in food-storing birds. *Philosophical Transactions - Royal Society of London, B*, *329*(1253), 143–151. https://doi.org/10.1098/rstb.1990.0159

Shettleworth, S. J. (2009). The evolution of comparative cognition: Is the snark still a boojum? *Behavioural Processes*, *80*(3), 210–217. https://doi.org/10.1016/J.BEPROC.2008.09.001

Shettleworth, S. J. (2010). Clever animals and killjoy explanations in comparative psychology. *Trends in Cognitive Sciences*, *14*(11), 477–481. https://doi.org/10.1016/j.tics.2010.07.002

Shimizu, Y. A., & Johnson, S. C. (2004). Infants' attribution of a goal to a morphologically unfamiliar agent. *Developmental Science*, *7*(4), 425–430. https://doi.org/10.1111/j.1467-7687.2004.00362.x

Siemann, M., & Delius, J. D. (1993). Implicit deductive responding in humans. *Naturwissenschaften*. https://doi.org/10.1007/BF01138792

Siemann, M., & Delius, J. D. (1998). Algebraic Learning and Neural Network Models for Transitive and Non-transitive Responding. In *European Journal of Cognitive Psychology* (Vol. 10). Retrieved from http://nbn-resolving.de/urn:nbn:de:bsz:352-205959

Siemann, M., Delius, J. D., & Wright, A. A. (1996). Transitive responding in pigeons: Influences of stimulus frequency and reinforcement history. *Behavioural Processes*, *37*(2–3), 185–195. https://doi.org/10.1016/0376-6357(96)00020-4

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*,

22(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Sirois, S., & Jackson, I. (2007). Social cognition in infancy: A critical review of research on higher order abilities. *European Journal of Developmental Psychology*, *4*(1), 46–64. https://doi.org/10.1080/17405620601047053

Smaldino, P. E. (2017). Models Are Stupid, and We Need More of Them. In *Computational Social Psychology* (pp. 311–331). https://doi.org/10.4324/9781315173726-14

Smulders, T. V, Sasson, A. D., & DeVoogd, T. J. (1995). Seasonal variation in hippocampal volume in a food-storing bird, the black-capped chickadee. *Journal of Neurobiology*, *27*(1), 15–25. https://doi.org/10.1002/neu.480270103

Sommerville, J. A., Woodward, A. L., & Needham, A. (2005). Action experience alters 3-month-old infants' perception of others' actions. *Cognition*, *96*(1), B1-11. https://doi.org/10.1016/j.cognition.2004.07.004

Southgate, V., Chevallier, C., & Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication. *Developmental Science*, *13*(6), 907–912. https://doi.org/10.1111/j.1467-7687.2009.00946.x

Southgate, V., Senju, A., & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science*, *18*(7), 587–592. https://doi.org/10.1111/j.1467-9280.2007.01944.x

Southgate, V., & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, *130*(1), 1–10. https://doi.org/10.1016/j.cognition.2013.08.008

Stamp Dawkins, M. (2002). What are birds looking at? Head movements and eye use in chickens. *Animal Behaviour*, *63*(5), 991–998. https://doi.org/10.1006/anbe.2002.1999

Steglich-Petersen, A., & Michael, J. (2015). Why Desire Reasoning is Developmentally Prior to Belief Reasoning. *Mind & Language*, *30*(5), 526–549. https://doi.org/10.1111/mila.12089

Stevens, J. R. (2017). Replicability and Reproducibility in Comparative Psychology. *Frontiers in Psychology*, *8*, 862. https://doi.org/10.3389/fpsyg.2017.00862

Surian, L., Caldi, S., & Sperber, D. (2007). Attribution of beliefs by 13-month-old infants. *Psychological Science*, *18*(7), 580–586. https://doi.org/10.1111/j.1467-9280.2007.01943.x

Teufel, C., Alexis, D. M., Clayton, N. S., & Davis, G. (2010). Mental-state attribution drives rapid, reflexive gaze following. *Attention, Perception, and Psychophysics*, *72*(3), 695–705. https://doi.org/10.3758/APP.72.3.695

Teufel, C., Clayton, N. S., & Russell, J. (2013). Two-Year-Old Children's Understanding of Visual Perception and Knowledge Formation in Others. *Journal of Cognition and Development*, *14*(2), 203–228. https://doi.org/10.1080/15248372.2012.664591

Tomasello, M., & Call, J. (2006). Do chimpanzees know what others see—or only what they are looking at? In *Rational Animals?* (pp. 371–384). https://doi.org/10.1093/acprof:oso/9780198528272.003.0017

Trafimow, D. (2012). The role of auxiliary assumptions for the validity of manipulations and measures. *Theory & Psychology*, *22*(4), 486–498. https://doi.org/10.1177/0959354311429996

Trezza, V., Campolongo, P., & Vanderschuren, L. J. M. J. (2011). Evaluating the rewarding nature of social

interactions in laboratory animals. *Developmental Cognitive Neuroscience*, *1*(4), 444–458. https://doi.org/10.1016/J.DCN.2011.05.007

van der Vaart, E., & Hemelrijk, C. K. (2012). 'Theory of mind' in animals: ways to make progress. *Synthese*, *191*(3), 335–354. https://doi.org/10.1007/s11229-012-0170-3

Van Elzakker, M., O'Reilly, R. C., & Rudy, J. W. (2003). Transitivity, flexibility, conjunctive representations, and the hippocampus. I. An empirical analysis. *Hippocampus*, *13*(3), 334–340. https://doi.org/10.1002/hipo.10083

Van Horik, J. O., Clayton, N. S., & Emery, N. J. (2012). Convergent Evolution of Cognition in Corvids, Apes and Other Animals. In *The Oxford Handbook of Comparative Evolutionary Psychology*. https://doi.org/10.1093/oxfordhb/9780199738182.013.0005

Vander Wall, S. B. (1990). *Food hoarding in animals*. University of Chicago Press.

Vasconcelos, M. (2008). Transitive inference in non-human animals: An empirical and theoretical analysis. *Behavioural Processes*, *78*, 313–334. https://doi.org/10.1016/j.beproc.2008.02.017

von Fersen, L., Wynne, C. D. L., Delius, J. D., & Staddon, J. E. R. (1991). Transitive Inference Formation in Pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*(3), 334–341. https://doi.org/10.1037/0097-7403.17.3.334

Vonk, J., & Povinelli, D. J. (2011). Social and Physical Reasoning in Human-reared Chimpanzees. In *Perception, Causation, and Objectivity* (pp. 342–368). https://doi.org/10.1093/acprof:oso/9780199692040.003.0019

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & Van Der Maas, H. L. J. (2011). *Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011)*. https://doi.org/10.1037/a0022790

Warneken, F., Hare, B., Melis, A. P., Hanus, D., & Tomasello, M. (2007). Spontaneous Altruism by Chimpanzees and Young Children. *PLoS Biology*, *5*(7), e184. https://doi.org/10.1371/journal.pbio.0050184

Warneken, F., & Tomasello, M. (2006). Altruistic Helping in Human Infants and Young Chimpanzees. *Science*, *311*(5765), 1301–1303. https://doi.org/10.1126/science.1121448

Watanabe, A., Grodzinski, U., & Clayton, N. S. (2014). Western scrub-jays allocate longer observation time to more valuable information. *Animal Cognition*, *17*(4), 859–867. https://doi.org/10.1007/s10071-013-0719-7

Weir, A. A. S., Chappell, J., & Kacelnik, A. (2002). Shaping of Hooks in New Caledonian Crows. *Science*, *297*(5583). Retrieved from http://science.sciencemag.org/content/297/5583/981

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition*, *30*(3), 239–277. https://doi.org/10.1016/0010-0277(88)90021-2

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: the truth about false belief. *Child Development*, *72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523–541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Wellman, H. M., & Woolley, J. D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition*, *35*(3), 245–275. https://doi.org/10.1016/0010-0277(90)90024-E

Whiten, A. (1996a). When does smart behaviour-reading become mind-reading? In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 277–292). https://doi.org/10.1017/CBO9780511597985.018

Whiten, A. (1996b). When does smart behaviour-reading become mind-reading? In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind*. https://doi.org/10.1017/CBO9780511597985

Whiten, A. (2013). Humans are not alone in computing how others see the world. *Animal Behaviour*, *86*(2), 213–221. https://doi.org/10.1016/j.anbehav.2013.04.021

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Woocher, F. D., Glass, A. L., & Holyoak, K. J. (1978). Positional discriminability in linear orderings. *Memory & Cognition*, *6*(2), 165–173. https://doi.org/10.3758/BF03197442

Woodward, A. L. (1998). Infants selectively encode the goal object of an actor's reach. *Cognition*, *69*(1), 1–34. https://doi.org/10.1016/S0010-0277(98)00058-4

Woodward, A. L. (1999). Infants' ability to distinguish between purposeful and non-purposeful behaviors. *Infant Behavior and Development*. https://doi.org/10.1016/S0163-6383(99)00007-7

Woodward, A. L. (2009). Infants' Grasp of Others' Intentions. *Current Directions in Psychological Science*, *18*(1), 53–57. https://doi.org/10.1111/j.1467-8721.2009.01605.x

Woodward, A. L., & Sommerville, J. A. (2000). Twelve-Month-Old Infants Interpret Action in Context. *Psychological Science*, *11*(1), 73–77. https://doi.org/10.1111/1467-9280.00218

Wynne, C. D. L. (1995). Reinforcement accounts for transitive inference performance. *Animal Learning & Behavior*, *23*(2), 207–217. https://doi.org/10.3758/BF03199936

Yott, J., & Poulin-Dubois, D. (2016). Are Infants' Theory-of-Mind Abilities Well Integrated? Implicit Understanding of Intentions, Desires, and Beliefs. *Journal of Cognition and Development*, *17*(5), 683–698. https://doi.org/10.1080/15248372.2015.1086771

Zentall, T. R. (2001). The case for a cognitive approach to animal learning and behavior. *Behavioural Processes*, *54*(1–3), 65–78. https://doi.org/10.1016/S0376-6357(01)00150-4

Appendix A – Bird Colonies

Table A1 Colony 1 – North Aviary

Name	M/F	Partner	Born	Died	Experiments[18]
Adlington	F	Pendleton	May 2008		5.2a, [5.2b]
Ainslie	M	N/A	May 2008	Sep 2017	
Ayton	M	Purchas	May 2008		4.1, (5.2a), 5.2b
Hoy	M	Hunter	May 2008		4.1, 4.2, [4.3], 5.1, 5.2a, 5.2b, 6.1
Hunter	F	Hoy	May 2008		(4.1), [4.2], 4.3, 5.1, 5.2a, 5.2b. [6.1]
Pendleton	M	Adlington	May 2008	Jan 2017	
Purchas	F	Ayton	May 2008		4.1, 5.1
Romero	M	Webb	May 2008		4.1, 4.2, [4.3], 5.1, 5.2a, 5.2b, 6.1
Webb	F	Romero	May 2008		4.1, [4.2], 4.3, (5.2a), [6.1]

Table A2 Colony 2 – Avian Building

Name	M/F	Partner	Born	Died	Experiments[19]
Caracas	M	Wellington	May 2006		(4.1), 4.2, (5.2a), 5.2b, 6.1
Dublin	M	Quito	May 2006	Jan 2019	(5.2a), 6.1
Jerusalem	F	Lima	May 2006	Jun 2018	4.1, [4.2], [6.1]
Lima	M	Jerusalem	May 2006		(4.1), 4.2, 5.1, 5.2a, 5.2b, 6.1
Lisbon	M	Rome	May 2006		(5.2a), (5.2b), 6.1
Quito	F	Dublin	May 2006		(5.2a), 5.2b, [6.1]
Rome	F	Lisbon	May 2006		[5.1], (5.2a), (5.2b), [6.1]
Washington	F	N/A	May 2006		5.2a, 5.2b
Wellington	F	Caracas	May 2006		4.1, [4.2], 5.2a, (5.2b), [6.1]

[18] Numbers in round brackets indicate an experiment in which the bird was either unable to complete testing or was excluded from the analysis based on pre-determined criteria. Numbers in square brackets indicate an experiment in which the bird acted only as a partner or observer of the focal bird.

Appendix B – Additional Bird Data

For access to the full raw data from the Eurasian jay experiments, as well as details about testing dates and training, visit https://osf.io/suz2a/?view_only=500398f1349446a5967ea4b62293a3ad.

Chapter 4: Experiment 1

Table B1 Looking times (in seconds) through the peephole into the open section and the peephole that into the closed section across the different trials. Three birds were excluded from testing best on pre-determined criteria; one because they failed the second familiarisation stage (Hunter), and two because their partners did not pass the first familiarisation stage (Caracas and Lima). It was predicted that the jays would look for longer through the peephole into the open section when their partner was able to enter (social condition). For the analysis the looking times were summed across the right and left trials within each condition.

	Non-social Right		Non-social Left		Social Right		Social Left	
	Open	Closed	Open	Closed	Open	Closed	Open	Closed
Romero	6.34	0	0	1.55	14.081	0	0	50.858
Hoy	4.82	0	0	0	29.162	0	0	12.149
Webb	19.729	0	4.344	2.48	3.889	3.059	0	30.859
Ayton	0	0.16	0	0	31.94	0	0	0.681
Purchas	0.859	2.399	5.918	0	0	0	0	31.001
Wellington	0	0	1.416	4.571	0	0	6.728	3.863
Jerusalem	2.46	1.414	4.309	0	0	12.288	8.6	6.357
Hunter	-	-	-	-	-	-	-	-
Caracas	-	-	-	-	-	-	-	-
Lima	-	-	-	-	-	-	-	-

Chapter 4: Experiment 2

Table B2 The number of boxes containing each food type opened by the males across the different conditions, W = waxworms, M = mealworms. In the baseline females were pre-fed maintenance diet. It was predicted that the males would show a greater preference for opening the boxes containing waxworms in the pre-fed mealworms trial than the pre-fed waxworms trial.

	Baseline		Pre-fed M		Pre-fed W	
	W	M	W	M	W	M
Caracas	10	9	6	8	8	3
Hoy	6	5	0	0	0	0
Lima	3	4	1	3	0	0
Romero	7	8	2	1	4	3

Table B3 The number of worms taken by the females from boxes opened by the males. The females showed a low rate of taking the worms from the opened boxes, despite being willing to take the food during familiarisation. This may have been due to the pre-feeding, which might have decreased their motivation to take the food as they were already sated.

	Baseline		Pre-fed M		Pre-fed W	
	W	M	W	M	W	M
Wellington	8	6	0	2	1	2
Hunter	5	5	0	0	0	0
Jerusalem	0	0	0	0	0	0
Webb	4	2	0	0	0	0

Chapter 5: Experiment 1

Table B4 The number of nuts cached in each trial, P = peanuts, M = macadamia nuts. The trials are labelled such that the actual food that the observer was pre-fed is on the left and the food the cacher saw entering the observer's is on the right i.e. 'actual/seen.' If the cacher was responding to the satiety of the observer they should show a greater preference for caching peanuts when the observer was pre-fed peanuts than when the observer was pre-fed macadamia nuts.

		Matching				Conflicting			
		P/P		M/M		M/P		P/M	
Cacher	Observer	P	M	P	M	P	M	P	M
Hoy	Purchas	20	0	16	0	30	0	23	0
Romero	Hunter	4	2	3	0	2	6	4	0
Purchas	Hoy	1	0	1	0	1	0	2	5
Hunter	Romero	0	1	1	0	0	2	0	1
Lima	Rome	10	22	15	15	0	14	30	6

Chapter 5: Experiment 2a

Table B5 Eurasian jay data from the pilfering experiment in the original study (Ostojić et al., 2017). The number of nuts pilfered in each condition, out of a total of 8 peanuts (P) and 8 macadamia nuts (M) cached by the human experimenter.

	Baseline		Pre-fed P		Pre-fed M	
	P	M	P	M	P	M
Lima	5	4	0	2	4	3
Caracas	2	3	1	1	2	1
Rome	0	1	0	0	0.5	0
Washington	0	1	0	1	2	0
Dublin	1	0	0	0	2	0
Jerusalem	0	1	0	0	1	1

Table B6 The number of nuts pilfered in each condition of my replication study, out of a total of 8 peanuts (P) and 8 macadamia nuts (M). Seven birds were tested but excluded based on pre-determined criteria; two were unable to complete the experiment for welfare reasons (Webb and Rome), one did not pass the baseline (Quito) and four did not pilfer any food in either of the two trials (Ayton, Caracas, Dublin, Lisbon). It was predicted that, as in the original study, pilferers would show a greater preference for pilfering peanuts when pre-fed macadamia nuts than when pre-fed peanuts.

	Baseline		Pre-fed P		Pre-fed M	
	P	M	P	M	P	M
Hoy	3	1	1	3	1	2
Washington	2	3	4	2	4	1
Romero	8	5	4	5	1	2
Hunter	1	1	1	1	2	2
Wellington	2	2	7	5	1	0
Adlington	0	1	2	0	0	0
Lima	6	6	5	4	3	6
Lisbon	4	3	0	0	0	0
Dublin	1	1	0	0	0	0
Caracas	1	0	0	0	0	0
Ayton	0	1	0	0	0	0
Webb	3	1	-	-	-	-
Quito	0	0	-	-	-	-
Rome	-	-	-	-	-	-

Chapter 5: Experiment 2b

Table B7 Eurasian jay data from the pilfering experiment in the original study (Ostojić et al., 2017). The number of nuts cached in each condition, P = peanuts, M = macadamia nuts. In the seen condition the cacher's could see the observer during pre-feeding, while in the unseen they could not.

		Seen						Unseen					
		Baseline		Pre-fed P		Pre-fed M		Baseline		Pre-fed P		Pre-fed M	
Cacher	Observer	P	M	P	M	P	M	P	M	P	M	P	M
Caracas	Quito	10	1	18	2	10	0	15	0	26	1	15.5	0
Wash.	Lima	1	3	1	1	2	10	2	1	0	0	4	3
Hoy	Pendle.	14	0	5	0	2	0	1	0	5	0	4	1
Rome	Quito	8	1	13	0	6	1	9	0	3	0	4	0
Dublin	Caracas	4.5	0	2	0	4	0	8	0	12	0	9	0
Romero	Adlington	3	0	2	0	2	0	2.5	0	3	0	0	0
Pendle.	Hoy	3	0	3	0	2	1	4	0	2	0	3	0

Table B8 The number of nuts cached in each condition of my replication study. It was predicted that, as in the original study, cachers would show a greater preference for caching peanuts when the observer was pre-fed peanuts than when the observer was pre-fed macadamia nuts.

		Seen						Unseen					
		Baseline		Pre-fed P		Pre-fed M		Baseline		Pre-fed P		Pre-fed M	
Cacher	Observer	P	M	P	M	P	M	P	M	P	M	P	M
Romero	Hoy	0	3	1	1	1	1	1	5	0	3	0	2
Ayton	Hoy	0	2	0	0	3	7	0	3	0	4	0	6
Hunter	Ayton	0	13	0	1	0	2	1	8	0	5	0	3
Hoy	Adlington	2	11	4	32	1	25	4	16	6	6	2	16
Quito	Well.	1	17	0	10	1	9	2	8	0	0	0	3
Lima	Wash.	5	0	14	1	10	16	3	32	8	26	0	29
Caracas	Lima	1	18	2	17	0	6	3	2	2	2	3	8
Wash.	Quito	1	2	3	4	4	3	2	3	1	4	0	5
Lisbon	Caracas	0	0	-	-	-	-	-	-	-	-	-	-
Rome	Caracas	-	-	-	-	-	-	0	2	0	0	0	0
Well.	Lima	-	-	-	-	-	-	0	0	-	-	-	-

Chapter 6: Experiment 1

Table B9 The number of worms shared by the male with their female partner in each of the trials, W = waxworms, M = mealworms. It was predicted that in the informative condition the male would share a greater proportion of waxworms when the female chose a waxworm than when the female chose a mealworm, but he would not necessarily do so in the uninformative condition.

			Uninformative				Informative			
	Baseline		M chosen		W chosen		M chosen		W chosen	
	W	M	W	M	W	M	W	M	W	M
Caracas	4	1	7	1	8	1	7	2	6	1
Lima	5	2	3	3	6	4	2	3	7	1
Dublin	5	6	4	4	3	3	0	3	5	1
Lisbon	4	5	2	3	5	4	3	3	4	1
Hoy	3	0	6	0	7	3	6	1	8	0
Romero	3	3	3	3	5	3	2	2	5	2

Appendix C –Human Data Summaries

For access to the full raw data from the human experiments, as well as examples of testing materials and images of the experiments, visit the OSF project for my PhD at: https://osf.io/suz2a/?view_only=500398f1349446a5967ea4b62293a3ad.

Chapter 2: Experiments 1-3

Table C1. Answers to the explicit question and looking times are given as mean ± standard deviation, with mean given to three significant figures and standard deviation to two. The first looks are given as the number of participants whose first looks were directed towards each bowl of food; A = apples, C = crackers, L = large amount of chocolate, S = small amount of chocolate. The proportion is calculated as A/(A+C) or L/(L+S).

	Explicit Question		First Look				Looking Times (s)					
			Apples		Crackers		Apples			Crackers		
Exp.	Apples	Crackers	A	C	A	C	A	C	Prop	A	C	Prop
1	4.95 ± 2.4	2.65 ± 2.4	11	14	11	14	0.699 ± 0.75	0.783 ± 0.93	0.465 ± 0.31	0.749 ± 0.71	0.655 ± 0.53	0.523 ± 0.32
2	5.09 ± 2.2	2.03 ± 2.2	9	16	19	6	0.763 ± 0.94	0.647 ± 0.47	0.454 ± 0.31	0.823 ± 0.57	0.648 ± 0.55	0.594 ± 0.31

	Match	Different	Match		Different		Match			Different		
			L	S	L	S	L	S	Prop	L	S	Prop
3	3.25 ± 2.5	4.70 ± 2.0	9	17	11	15	0.658 ± 0.34	0.791 ± 0.75	0.511 ± 0.26	0.654 ± 0.41	0.646 ± 0.52	0.516 ± 0.27

Chapter 3: Experiments 1 and 2

Table C2 Participants' answers to the intentional action question are given as the number and percentage of participants who chose to open the correct door. The ranking questions show the mean rank each object was given and the percentage of participants who put the object in the correct position in the hierarchy.

	Intentional action			Ranking questions					
Exp.				A	B	C	D	E	All
1	Correct	146	Mean rank	2.87	2.62	2.76	2.71	4.05	3.00
	Percentage	50.0%	Percentage Correct	24.3	24.3	20.9	18.5	50.0	3.77
2a	Correct	33	Mean rank	1.55	1.73	2.71			2.00
	Percentage	67.3%	Percentage Correct	51.0	44.9	79.6			40.8
2b	Correct	27	Mean rank	1.88	1.78	2.34			2.00
	Percentage	54.0%	Percentage Correct	42.0	38.0	52.0			22.0

Chapter 6: Experiment 2

Table C3 Participant's reaction times are given for each trial as mean ± standard deviation, with mean to three significant figures and standard deviation to two significant figures. In the pilot, Experiment 2a, all trials were informative with two congruent trials (1C and 2C) and two incongruent trials (1I and 2I). In Experiment 2b the first letter indicates whether the trial was informative (I) or uninformative (U), and the second letter indicates whether the trial was congruent (C) or incongruent (I). The reaction times were calculated from the start of video, which included a one second delay before the agent started to move. The differences were calculated as incongruent – congruent. If the participants were sensitive to the agent's preference it was predicted that they would respond more slowly in the incongruent trials than the congruent trials.

	Reaction times (s)				Difference	
Exp.	1C	2C	1I	2I	First pair	Second pair
2a	1.62 ± 0.10	1.65 ± 0.14	1.66 ± 0.13	1.61 ± 0.091	0.0480 ± 0.082	-0.0395 ± 0.10

	IC	II	UC	UI	Informative	Uninformative
2b	1.56 ± 0.35	1.55 ± 0.33	1.61 ± 0.21	1.66 ± 0.29	-0.0112 ± 0.21	0.0453 ± 0.19